arXiv:1502.02291v1 [math.ST] 8 Feb 2015

Regularized Functional Canonical Correlation Analysis for Stochastic Processes

David B. King

405 N. Park Ave., Indiana University, Bloomington, IN 47408

Abstract

In this paper we derive the asymptotic distributions of two distinct regularized estimators for functional canonical correlation as well as their associated eigenvalues, eigenvectors and projection operators. The methods we developed utilize regularized estimators which approach the functional operators based in reproducing kernel Hilbert spaces (RKHS) as the regularization parameter approaches zero. In addition to providing some justification for the RKHS methods, we explore the asymptotics of regularized operators associated with both Tikhinov and truncated singular value decomposition (TSVD) type regularization. Together, these regularization methods represent two of the most commonly utilized forms of regularization.

Keywords: Canonical Correlation; Asymptotic Distributions; Stochastic Processes; Reproducing Kernel Hilbert Spaces; Regularization; Inverse Problems
AMS 2000 Subject Classification: Primary 62H20, 60E05, 62H25, 62M99, 45B05, 45Q05

1. Introduction

The goal of multivariate canonical correlation analysis (MCCA; Hotelling [19]) is to identify and quantify the associations between two random vectors $\mathbf{X}_1 \in \mathbf{R}^{N_1}$ and $\mathbf{X}_2 \in \mathbf{R}^{N_2}$. Recently interest has been focused on the extension of this notion to the collection and analysis of "functional data" where the term refers to observations that are curves or sample paths of continuous time stochastic processes. Although development of statistical methodology for the analysis of functional data has been an active research area for well over twenty years, the current popularity of functional data analysis (FDA) is due, in large part, to monographs by Ramsay and Silverman [30] [29]. What separates functional data from ordinary multivariate data is that the observed data are sample paths from stochastic processes $X_1(\cdot)$ and $X_2(\cdot)$, which are assumed to be elements of some infinite dimensional and separable Hilbert space

Email address: dbking@indiana.edu (David B. King)

consisting of functions defined on an index set E, such as $[0, 1]$ or $\mathbb{Z}$. In this setting, the covariance matrices which are central to the development of the theory of MCCA are replaced by covariance operators of integral type. In this infinite dimensional case, some difficulties regarding the definition of the sample canonical correlation have already been observed in Leurgans et al. [25]. These authors argue that some kind of smoothing or regularization is indispensable when dealing with estimating the sample canonical correlation. The source of the difficulty in the functional data case is that the sample estimators for covariance operators have finite rank while, in principal, they operate in an infinite dimensional Hilbert space. Leurgans et al. [25] points out that as a consequence, the sample principal canonical correlation will always be 1 if no regularization or smoothing is done. This problem originates from the fact that when the number of time points at which the processes are measured becomes larger than the sample size, it will always be possible to find linear combinations of both processes which are perfectly correlated. From a functional analysis standpoint, the covariance operators involved in the analysis require regularization as they are Hilbert-Schmidt and thus do not possess an inverse (see e.g., Rynne and Youngson [33]). The situation involved with functional canonical correlation analysis (FCCA) is analogous, therefore, to the classic inverse problem of finding approximate solutions to equations involving Freidholm integral equations. Much like this classic problem, regularization plays an instrumental role in it's resolution.

This paper is organized as follows. In Section 2 we introduce the notations, definitions and assumptions which we will utilize throughout the paper. In this section we will also discuss why reproducing kernel Hilbert space (RKHS) methods is the ideal Hilbert space to solve the functional canonical correlation analysis (FCCA) problem. In Section 3 we will introduce the notions of canonical correlation and discuss why the Eubank and Hsing [15] approach to FCCA provides most complete definition to canonical correlation analysis without regularization. In Section 4 we will discuss the general theory associated with regularization and introduce both the Tikhinov and truncated singular value decomposition (TSVD) types of regularization (Engl et al. [14]). In Section 5 and 6 we will discuss the consistency and asymptotic distributional theory associated with Tikhinov regularized canonical correlation operators, and in Sections 7 and 8 we will do the same with TSVD regularization. Finally, Section 9 will be devoted to summarizing our conclusions and providing some further recommendations.

2. Basic notation, definitions and assumptions

Let E be a subset of $\mathbb{R}$ and ν a sigma-finite measure on E. We then consider the case where a stochastic process $\{X(t), t \in E\}$ takes values in the Hilbert space $\mathcal{H} = L^2(E)$ of square integrable functions on E with inner product $\langle f, g\rangle_{\mathcal{H}} \equiv \int_E f(t)g(t)d\nu(t)$. Throughout it will be assumed that

$$\mathrm{E}\|X\|^4_{\mathcal{H}} < \infty. \tag{1}$$

Under this assumption, $\mathrm{E}\langle X, f\rangle_{\mathcal{H}} = \int_{\mathcal{H}} \langle x, f\rangle_{\mathcal{H}} dP(x) < \infty$ for all $f \in \mathcal{H}$ with P denoting the induced probability measure of X on $\mathcal{H}$. The Riesz-Frechet representation theorem then ensures the existence of an element $\mu \in \mathcal{H}$ such that $\mathrm{E}\langle X, f\rangle = \langle \mu, f\rangle$. Under assumption (1), the Riesz-Frechet representation theorem also ensures the existence of the covariance operator $S : \mathcal{H} \mapsto \mathcal{H}$, which is given by

$$\mathrm{E}[\langle f, X-\mu\rangle\langle X-\mu, g\rangle] = \mathrm{E}[\langle f, (X-\mu) \otimes_{\mathcal{H}} (X-\mu)g\rangle] = \langle f, Sg\rangle \qquad (2)$$

where $\otimes_{\mathcal{H}}$ is the tensor product in $\mathcal{H}$ and is defined by $(f \otimes_{\mathcal{H}} g)h \equiv \langle f, h\rangle_{\mathcal{H}} g$ for all $f, g, h \in \mathcal{H}$. We may also write $S = \mathrm{E}[(X-\mu) \otimes_{\mathcal{H}} (X-\mu)]$. It is also well known that the covariance operator S is self-adjoint, non-negative definite, and has finite trace (see Laha and Rohatgi [33]). The finite trace property ensures that S is Hilbert-Schmidt and hence compact.

For any abstract Hilbert spaces $\mathcal{M}$ and $\mathcal{N}$ let $\mathcal{B}(\mathcal{M}, \mathcal{N})$ denote the Banach space of all bounded operators that map $\mathcal{M}$ to $\mathcal{N}$. A subclass of $\mathcal{B}(\mathcal{M}, \mathcal{N})$ is $\mathcal{K}(\mathcal{M}, \mathcal{N})$ which will denote the set of all compact operators that map $\mathcal{M}$ to $\mathcal{N}$. Of particular importance in this paper is the subclass of compact operators which have finite trace, known as Hilbert-Schmidt operators. Let $\mathcal{K}_{HS}(\mathcal{M}, \mathcal{N})$ denote the set of all Hilbert-Schmidt operators that map $\mathcal{M}$ to $\mathcal{N}$. In this paper we will use the simplifying notation that $\mathcal{B}(\mathcal{M}) = \mathcal{B}(\mathcal{M}, \mathcal{M})$, $\mathcal{K}(\mathcal{M}) = \mathcal{K}(\mathcal{M}, \mathcal{M})$ and $\mathcal{K}_{HS}(\mathcal{M}) = \mathcal{K}_{HS}(\mathcal{M}, \mathcal{M})$. The ordinary operator norm on $\mathcal{B}(\mathcal{M})$ will be denoted by $\|\cdot\|$. The set of Hilbert-Schmidt operators $\mathcal{K}_{HS}(\mathcal{M})$ becomes a separable Hilbert space when it is endowed with the inner product

$$\langle A, B\rangle_{HS} = \sum_{k=1}^{\infty} \langle Ae_k, Be_k\rangle_{\mathcal{M}} = \mathrm{tr}(A^*B), \quad A, B \in \mathcal{K}_{HS}(\mathcal{M}) \qquad (3)$$

with $\{e_k\}_{k=1}^{\infty}$ denoting any complete orthonormal system (CONS) for $\mathcal{M}$. This inner product does not depend on the choice of basis (Kato [22]). The inner product, norm and tensor product on $\mathcal{K}_{HS}(\mathcal{M})$ will be denoted by $\langle\cdot,\cdot\rangle_{HS}$, $\|\cdot\|_{HS}$ and $\otimes_{HS}$ respectively.

Next, assume that $\mathcal{H}_1$ and $\mathcal{H}_2$ are two closed subspaces of $\mathcal{H}$ such that

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2, \quad \mathcal{H}_1 \perp \mathcal{H}_2$$

and let Υ_i, $i = 1, 2$ denote the orthogonal projection operator of $\mathcal{H}$ onto $\mathcal{H}_i$ for $i = 1, 2$. Suppose further that $X_i = \Upsilon_i X$, $\mu_i = \Upsilon_i \mu$, and S_{ij} denote the restriction of S to $\mathcal{H}_i$ and $\mathcal{H}_j$ for $i, j = 1, 2$ so that $S_{ij} = \Upsilon_j S \Upsilon_i$. Because the Υ_i are bounded and S is Hilbert-Schmidt, the S_{ij} for $i, j = 1, 2$ are also Hilbert-Schmidt and compact. In addition, the S_{ii} are self-adjoint and non-negative definite. For convenience, we henceforth denote $S_{ii} = S_i$.

For $i = 1, 2$, let $\{\phi_{in}\}_{n=1}^{\infty}$ be an orthonormal basis corresponding to eigenvectors of S_i with $\{\lambda_{in}\}_{n=1}^{\infty}$, the corresponding sequence of non-negative eigenvalues. Since S_i is self-adjoint, non-negative and compact we may write

$$S_i = \sum_{n=1}^{\infty} \lambda_{in} \phi_{in} \otimes_{\mathcal{H}_i} \phi_{in}, \quad i = 1, 2 \qquad (4)$$

with $\lambda_{i1} \geq \lambda_{i2} \geq \cdots \geq 0$ a decreasing sequence whose only limit can be zero. For our purposes we might as well assume without loss of generality (WLOG) that $\{\phi_{in}\}_{n=1}^{\infty}$ is a CONS for $\mathcal{H}_i$, S_i is strictly positive and $\mathcal{H}_i = \ker(S_i)^{\perp}$ for $i = 1, 2$. We make this assumption since if $\varphi \in \ker(S_i)$ then $\mathrm{Var}[\langle \varphi, X_i \rangle_{\mathcal{H}}] = \langle \varphi, S_i \varphi \rangle_{\mathcal{H}} = 0$, which would have the consequence that $\langle \varphi, X_i \rangle_{\mathcal{H}_i} = \langle \mu_i, \varphi \rangle_{\mathcal{H}_i}$ with probability one. It is also convenient at this juncture to assume, WLOG, that the mean of the process is zero because if this does not hold we may always consider the covariance of the process $X(\cdot) - \mu(\cdot)$ instead. It should be mentioned that in (4) the list of eigenvalues $\{\lambda_{in}\}$ is repeated according to their multiplicity. An alternative expression for (4) involving eigenprojection operators is

$$S_i = \sum_{h=1}^{\infty} \tilde{\lambda}_{ih} P_{ih}, \quad \text{for } i = 1, 2 \tag{5}$$

where $\{\tilde{\lambda}_{ih}\}$ are the distinct elements of $\{\lambda_{in}\}$, and P_{ih} is the finite dimensional projection operator onto the eigenspace associated with each distinct $\tilde{\lambda}_{ih}$ given by

$$P_{ih} = \sum_{\phi_{in}:\ \lambda_{in} = \tilde{\lambda}_{ih}} \phi_{in} \otimes_{\mathcal{H}_i} \phi_{in}. \tag{6}$$

Since the processes $\{X_i(\cdot)\}_{i=1}^{2}$ are of second order, they admit a Karhunen–Loève expansion $X_i(\cdot) = \sum_{n=1}^{\infty} Z_{in} \phi_{in}(\cdot)$ with the random variables Z_{in} defined by $Z_{in} = \langle X_i, \phi_{in} \rangle_{\mathcal{H}_i}$ (see Ash and Gardiner [3] or Doob [12]). These variables are orthogonal in the sense that $\mathrm{Cov}[Z_{ij}, Z_{ik}] = \langle \phi_{ij}, S_i \phi_{ik} \rangle_{\mathcal{H}_i} = \lambda_{ij} \delta_{jk}$ with δ_{jk} denoting the Kronecker delta function. Mercer's theorem then ensures that the covariance functions of the processes $\{X_i(t), t \in E_i\}_{i=1}^{2}$ are

$$\begin{aligned} K_{ii}(s,t) = \mathrm{E}[X_i(s) X_i(t)] &= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \mathrm{E}[Z_{im} Z_{in}] \phi_{in}(s) \phi_{im}(t) \\ &= \sum_{n=1}^{\infty} \lambda_{in} \phi_{in}(s) \phi_{in}(t) \quad \text{for } i = 1, 2. \end{aligned} \tag{7}$$

Moreover, the cross-covariance kernel is then

$$\begin{aligned} K_{12}(s,t) = \mathrm{E}[X_1(s) X_2(t)] &= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \mathrm{E}[Z_{1n} Z_{2m}] \phi_{1n}(s) \phi_{2m}(t) \\ &= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \gamma_{mn} \phi_{1n}(s) \phi_{2m}(t) \end{aligned} \tag{8}$$

and we note that $K_{12}(s,t) = K_{21}(s,t)$. For notational simplicity let $K_{ii}(s,t) = K_i(s,t)$. It is also well known that for all $f \in \mathcal{H}_i$

$$(S_{ij} f)(t) = \int_E K_{ij}(s,t) f(s) d\nu(s) = \langle K_{ij}(\cdot, t), f \rangle_{\mathcal{H}_i} \quad \text{for } i, j = 1, 2. \tag{9}$$

An alternate form for the cross-covariance operator $S_{12} : \mathcal{H}_2 \mapsto \mathcal{H}_1$ is given by

$$S_{12} = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \gamma_{mn} \phi_{2n}(s) \otimes_{\mathcal{H}_2} \phi_{1m}(t) = S_{21}^*. \tag{10}$$

In addition to $\mathcal{H} = \ker(S)^{\perp} \subseteq L^2(E)$, two additional types of Hilbert spaces will play prominent roles in further developments. The first type of Hilbert space are the reproducing kernel Hilbert spaces (RKHS) associated with the symmetric covariance kernels $K_i(\cdot, *)$, denoted $\mathcal{H}(K_i)$ (see Aronszajn [2] or Berlinet and Thomas-Agnan [4]). The second type are the Hilbert spaces generated by each stochastic process, denoted $L^2_{X_i}$ with $i = 1, 2$ (see Parzen [28]). To construct both of these Hilbert spaces we first let $\{t_1, \dots, t_n\}$ be any finite collection of points in E and let $\mathbf{X}_{in} = [X_i(t_1), \dots, X_i(t_n)]'$ with $\mathbf{K}_{in} = \{K_i(t_j, t_k)\}_{j,k=1}^n$ denoting the covariance matrix of $\mathbf{X}_{in}$ for each $n \in \mathbb{N}$. Next we define the pre-Hilbert space generated by the process to be the set of all arbitrary finite dimensional linear combinations of the process, i.e. $L^2_{X_{in}} = \{U = \mathbf{a}'\mathbf{X}_{in} : \mathbf{a} \in \ker(\mathbf{K}_{in})^{\perp} \subseteq \mathbb{R}^n\}$ where the inner product between two elements is given by

$$\langle \mathbf{a}'\mathbf{X}_{in}, \mathbf{b}'\mathbf{X}_{in} \rangle_{L^2_{X_i}} = \mathrm{Cov}[\mathbf{a}'\mathbf{X}_{in}, \mathbf{b}'\mathbf{X}_{in}] = \mathbf{a}'\mathbf{K}_{in}\mathbf{b}. \tag{11}$$

Likewise, the pre-Hilbert space of the RKHS is defined to be the column space of $\mathbf{K}_{in}$, i.e. $\mathcal{H}(\mathbf{K}_{in}) = \{\mathbf{f} = \mathbf{K}_{in}\mathbf{a} : \mathbf{a} \in \ker(\mathbf{K}_{in})^{\perp} \subseteq \mathbb{R}^n\}$ and the inner product between any $\mathbf{f} = \mathbf{K}_{in}\mathbf{a}$ and $\mathbf{g} = \mathbf{K}_{in}\mathbf{b}$ is given by

$$\langle \mathbf{f}, \mathbf{g} \rangle_{\mathcal{H}(\mathbf{K}_{in})} \equiv \mathbf{f}'\mathbf{K}_{in}^{\dagger}\mathbf{g} = \mathbf{a}'\mathbf{K}_{in}\mathbf{b} \tag{12}$$

where $\mathbf{K}_{in}^{\dagger}$ denotes the Moore-Penrose inverse of $\mathbf{K}_{in}$. The Parzen–Loéve congruence mapping is determined uniquely by $\Psi_i(K_i(t, \cdot)) = X_i(t)$ for each $t \in E$ with the result that every linear combination U of the $\mathbf{X}$ vector with nonzero variance can be expressed as

$$U = \Psi(\mathbf{f}) = \mathbf{f}'\mathbf{K}_{in}\mathbf{X}_{in} \tag{13}$$

for some $\mathbf{f} \in \mathcal{H}(\mathbf{K}_{in})$ (see King [20]). It is a simple matter to see that inner product given by (12) satisfies the reproducing property (see Aronszajn [2]). To see this let k be any index in $1, \dots, n$ and let $\mathbf{K}_{in}(t_k, \cdot) = \mathbf{K}_{in}^T(\cdot, t_k)$ denote the k^{th} row of $\mathbf{K}_n$. Now for any $\mathbf{f} = \mathbf{K}_{in}\mathbf{a} \in \mathcal{H}(\mathbf{K}_{in})$ we have that

$$\langle \mathbf{K}_n(\cdot, t_k), \mathbf{f} \rangle_{\mathcal{H}(\mathbf{K}_{in})} = \mathbf{K}_{in}(t_k, \cdot)\mathbf{K}_{in}^{\dagger}\mathbf{K}_{in}\mathbf{a} = \mathbf{K}_{in}(t_k, \cdot)\mathbf{a} = \mathbf{f}(t_k).$$

This demonstrates that the pre-Hilbert space $\mathcal{H}(\mathbf{K}_n)$ given by inner product defined in (12) must be the unique RKHS of the process $\{X(t_i)\}_{i=1}^n$. To complete the construction of $\mathcal{H}(K_i)$ and $L^2_{X_i}$ we then extend the realm of the pre-Hilbert spaces, which presently apply to any finite collection of points $\{t_1, \dots, t_n\} \in E$ to the index set, E in its entirety. This construction is accomplished through Cauchy completion or adding in the limits of arbitrary linear combinations of the form $\sum_{j=1}^n a_i K(\cdot, t_j)$ and $\sum_{j=1}^n a_i X(t_j)$. In this fashion, we see that

$$\mathcal{H}(K_i) = \overline{\{f : f(\cdot) = \int_E K(\cdot, t)a(t)d\nu(t)\}} = \overline{\mathrm{Im}(S_i)} = \ker(S_i)^{\perp} \tag{14}$$

and

$$L^2_{X_i} = \overline{\{U : U = \int_E a(t)X(t)d\nu(t)\}} \tag{15}$$

with $\overline{A}$, denoting the closure of any set A. In this infinite dimensional setting the RKHS is the set of function on E given by

$$\mathcal{H}(K_i) = \{f : f(\cdot) = \sum_{j=1}^{\infty} \lambda_{ij} f_{ij} \phi_{ij}(\cdot), \|f\|^2_{\mathcal{H}(K_i)} = \sum_{j=1}^{\infty} \lambda_{ij} f^2_{ij} < \infty\} \tag{16}$$

where $f_{ij} = \langle f, \phi_{ij}\rangle_{\mathcal{H}_{ij}}$ are the generalized Fourier coefficients relative to the CONS $\{\phi_{ij}\}_{j=1}^{\infty}$ for $\mathcal{H}_i$, $i = 1, 2$. An application of the integral representation theorem of Parzen [28] then produces the following result.

Theorem 2.1. *For $i = 1, 2$ let $f(\cdot) = \sum_{j=1}^{\infty} \lambda_{ij} f_{ij} \phi_{ij}(\cdot)$ be in $\mathcal{H}(K_i)$. Then,*

$$\Psi_i(f) = \sum_{j=1}^{\infty} f_{ij} Z_{ij} \text{ and } \Psi_i^{-1}\left(\sum_{j=1}^{\infty} f_{ij} Z_{ij}\right) = \sum_{j=1}^{\infty} \lambda_{ij} f_{ij} \phi_{ij} \tag{17}$$

with $Z_{ij} = \langle X_i, \phi_{ij}\rangle_{\mathcal{H}_i}$ and $\Psi_i^{-1} = \Psi_i^*$, where Ψ_i^* denotes the adjoint of Ψ_i.

The importance of the RKHS inner product when formulating theory regarding integral operators was shown by Nasheed and Wahba [27]. These authors provided a characterization of the RKHS $\mathcal{H}(K_i)$ generated by the kernel K_i to closure of the image of the integral operator for the symmetric square root, $\overline{\mathrm{Im}(S_i^{1/2})}$. In this regard, first notice that since the S_i are positive (and self-adjoint), they have symmetric square roots $S_i^{1/2}$ with associated symmetric kernel $\Phi_i(s, t)$ given explicitly by

$$\Phi_i(s, t) = \sum_{j=1}^{\infty} \lambda_{ij}^{1/2} \phi_{ij}(s)\phi_{ij}(t), \quad i = 1, 2. \tag{18}$$

For $i = 1, 2$ the symmetric kernels $\Phi_i(s, t)$ satisfy

$$K_i(s, t) = \int_E \Phi_i(s, r)\Phi_i(t, r)d\nu(r). \tag{19}$$

We further note that

$$\mathrm{Im}(S_i^{1/2}) \subseteq \overline{\mathrm{Im}(S_i^{1/2})} = \ker(S_i^{1/2})^{\perp} = \ker(S_i)^{\perp} = \mathcal{H}_i.$$

Nasheed and Wahba [27] then arrive at the following important theorem.

Theorem 2.2. *(Nasheed and Wahba, 1974) For $i = 1, 2$ the RKHS $\mathcal{H}(K_i)$ consist of functions of the form*

$$f(\cdot) = \int_E g(s)\Phi_i(\cdot, s)d\nu(s)$$

for some $g \in \mathcal{H}_i$. The inner product in $\mathcal{H}(K_i)$ is

$$\langle f_1, f_2 \rangle_{\mathcal{H}(K_i)} = \langle g_1, g_2 \rangle_{\mathcal{H}_i} \tag{20}$$

where $g_1, g_2 \in \mathcal{H}_i$ are the minimal $L^2(E)$ norm solutions of

$$f_j(\cdot) = \int_E g_j(s)\Phi_i(\cdot, s)d\nu(s),\ j = 1, 2.$$

Proof: For $i = 1, 2$, let V_i be the smallest closed subspace of $\mathcal{H}_i$ that contains $\Phi_i(\cdot, t)$ for all $t \in E$. Since the smallest linear space containing $\Phi_i(\cdot, t)$ for all $t \in E$ is $\text{span}\{\Phi_i(\cdot, t) : t \in E\}$, it follows that $V = \overline{\text{span}}\{\Phi_i(\cdot, t) : t \in E\}$. Now the projection theorem ensures that for each $f \in \mathcal{H}(K)$ there exists a unique element $g_f \in V$ of minimal norm which is the best-approximate solution to the inverse problem

$$f(t) = (S_i^{1/2} g_f)(t) = \int_E g_f(s)\Phi_i(s, t)d\nu(s),\ \forall t \in E.$$

Because g_f is unique, the inner product given by (20) and associated norm are well defined. We now only need to show that $K_i(\cdot, \cdot)$ are the reproducing kernel. However,

$$K_i(\cdot, t) = \langle \Phi_i(*, \cdot), \Phi i(*, t) \rangle_{\mathcal{H}_i}.$$

Thus, by (20),

$$\langle K_i(\cdot, t), f \rangle_{\mathcal{H}(K_i)} = \langle \Phi_i(*, t), g_f(*) \rangle_{\mathcal{H}_i} = (S_i^{1/2} g_f)(t) = f(t).$$

$\diamondsuit$

This theorem shows that for $i = 1, 2$ the optimal Hilbert space to solve inverse problems associated with integral equations of the form $f = S_i^{1/2} g$ is in the RKHS setting $\mathcal{H}(K_i)$. To illustrate this, consider the problem of finding a function $g(\cdot)$ to satisfy $S_i^{1/2} g = \int_E g(s)\Phi_i(\cdot, s)d\nu(s) = f(\cdot)$ for some given $f(\cdot) \in \mathcal{H}_i$. A least-squares solution to this problem is a minimizer of $\|S_i^{1/2} g - f\|_{\mathcal{H}_i}$ and a best least-squares solution is the one with minimum norm. If we let $F_i = (\text{Im}S_i^{1/2}) \oplus (\text{Im}S_i^{1/2})^{\perp}$ and assume $f \in F_i$, g is a least squares solution if and only if $S_i^{1/2} S_i^{1/2} g = S_i^{1/2} f = Sg$. Furthermore, the unique best least-squares solution is given by $g = (S_i)^{\dagger} S_i^{1/2} f = S_i^{1/2\dagger} f$ with $S_i^{1/2\dagger}$ denoting the Moore-Penrose inverse of $S_i^{1/2}$, and no least-squares solution exists if $f \notin F_i$. However, from Engl et al. [14], $f \in F_i$ if and only if f satisfies the Picard criterion

$$\sum_{j=1}^{\infty} \frac{\langle f, \phi_{ij} \rangle^2_{\mathcal{H}_i}}{\lambda_{ij}} < \infty \tag{21}$$

and, in that case,

$$g(\cdot) = (S_i^{1/2\dagger} f)(\cdot) = \sum_{j=1}^{\infty} \frac{\langle f, \phi_{ij} \rangle_{\mathcal{H}_i}}{\sqrt{\lambda_{ij}}} \phi_{ij}(\cdot).$$

Note that $f \in F_i$ if and only if $\|f\|_{\mathcal{H}(K_i)} = \sum_{j=1}^{\infty} \frac{\langle f, \phi_{ij} \rangle^2_{\mathcal{H}_i}}{\lambda_{ij}} < \infty$. Consequently, $\mathcal{H}(K_i) = \overline{(\mathrm{Im} S_i^{1/2})} = \ker(S_i)^{\perp}$, under the inner product

$$\langle f_{i1}, f_{i2} \rangle_{\mathcal{H}(K_i)} = \langle S_i^{1/2\dagger} f_{i1}, S_i^{1/2\dagger} f_{i2} \rangle_{\mathcal{H}_i}$$

with

$$S_i^{1/2\dagger} f_{ij}(\cdot) = \sum_{k=1}^{\infty} \frac{\langle f_{ij}, \phi_{ik} \rangle_{\mathcal{H}_i}}{\sqrt{\lambda_{ik}}} \phi_i(\cdot)$$

for $i, j = 1, 2$.

For further developments, a congruence which connects $\mathcal{H}_i$ to $\mathcal{H}(K_i)$ must be established.

Corollary 2.1. *(Eubank and Hsing, 2008) For $i = 1, 2$ the Hilbert spaces $\overline{\mathrm{Im}(S_i^{1/2})} = \ker(S_i)^{\perp}$ and $\mathcal{H}(K_i)$ are congruent under the mapping $\Gamma_i : \mathcal{H}_i \mapsto \mathcal{H}(K_i)$ defined by*

$$(\Gamma_i g)(\cdot) \equiv \sum_{j=1}^{\infty} \sqrt{\lambda_{ij}} g_{ij} \phi_{ij}(\cdot) \tag{22}$$

where $g = \sum_{j=1}^{\infty} \langle g, \phi_{ij} \rangle_{\mathcal{H}_i} \phi_{ij} = \sum_{j=1}^{\infty} g_{ij} \phi_{ij} \in \ker(S)^{\perp}$. The inverse mapping

$$(\Gamma_i^{-1} f)(\cdot) \equiv \sum_{j=1}^{\infty} \sqrt{\lambda_{ij}} f_{ij} \phi_{ij}(\cdot) \tag{23}$$

for $f = \sum_{j=1}^{\infty} \lambda_{ij} f_{ij} \phi_{ij}(\cdot) \in \mathcal{H}(K_i)$, is also the adjoint of Γ_i.

Note that for $i = 1, 2$ the operators Γ_i and $S_i^{1/2}$ are equal in the sense that for any $f \in \mathcal{H}_i$, $\Gamma f = \sum_{j=1}^{\infty} \sqrt{\lambda_{ij}} f_{ij} \phi_{ij} = S^{1/2} f$. The difference is in terms of the norm and inner product for the range of each operator.

3. Canonical Correlation

The literature on functional canonical correlation can be roughly dichotomized into formulations involving Hilbert space valued processes in $\mathcal{H}_i = \ker(S_i)^{\perp}$ (see He et al. [17] [18]) and an alternative approach that relies on reproducing kernel Hilbert space (RKHS) theory (Eubank and Hsing, [15]). In this section we will compare and contrast these two different approaches to functional CCA. In the He et al. [17] approach the k^{th} squared canonical correlation ρ_k^2 and associated weight functions f_k and g_k are found by the singular value decomposition of the cross-correlation operator of X_1 and X_2 defined by

$$R = S_1^{1/2\dagger} S_{12} S_2^{1/2\dagger} \tag{24}$$

where $S_i^{1/2\dagger}$ denotes the Moore-Penrose generalized inverse of $S_i^{1/2}$ for $i = 1, 2$ and is given explicitly by

$$S_i^{1/2\dagger} = \sum_{h=1}^{\infty} \tilde{\lambda}_{ih}^{-1/2} P_{ih}. \tag{25}$$

The right hand and left hand eigenvectors are then found by eigenvalue and eigenvector analysis of the operators RR^* and R^*R. The basic problem is that unlike the usual situation in the finite-dimensional case, the square roots of covariance operators of infinite dimensional Hilbert space valued processes are not invertible. To resolve this issue, He et al. [17] restricts the domain of $\{\mathcal{H}_1, \mathcal{H}_2\}$ to the subspace where the Moore-Penrose inverses of $S_1^{1/2}$ and $S_2^{1/2}$ can be defined. Thus, for $i = 1, 2$, the domain of $S_i^{1/2\dagger}$ is restricted to $F_i \equiv \{S_i^{1/2} h : h \in \ker(S_i)^\perp\}$ and is characterized as the set of functions satisfying the Picard criterion (21) (see Engl et al. [14]). Now, subject to the restriction that the domain of R be F_2, let $\rho_1^2 \geq \rho_2^2 \geq \cdots \geq 0$ denote the eigenvalues of R^*R with $g_1, g_2, \ldots \in F_2$ the corresponding eigenvectors. The left hand eigenvectors are obtained by $f_k = Rg_k/\rho_k \in F_1$. The canonical correlations and weight functions are $\{\rho_k, u_k = S_1^{1/2\dagger} f_k, v_k = S_2^{1/2\dagger} g_k\}_{k=1}^\infty$ and the corresponding canonical variables are $\{U_k = \langle u_k, X_1\rangle_{\mathcal{H}_1}, V_k = \langle v_k, X_2\rangle_{\mathcal{H}_2}\}_{k=1}^\infty$. In the He et al. [17] method the weight functions are not well defined whenever $f_k \notin F_1$ or $g_k \notin F_2$.

In contrast to the He et al. [17] method, the approach of Eubank and Hsing [15] involves the singular value decomposition of the RKHS based operator $T : \mathcal{H}(K_2) \mapsto \mathcal{H}(K_1)$ defined such that for any $\tilde{g} \in \mathcal{H}(K_2)$

$$(T\tilde{g})(s) = \langle K_{12}(s, \cdot), \tilde{g}\rangle_{\mathcal{H}(K_2)}. \tag{26}$$

Let $\{\rho_k^2, \tilde{g}_k\}_{k=1}^\infty$ denote the eigenvalues and eigenvectors of T^*T and $\{\rho_k^2, \tilde{f}_k\}_{k=1}^\infty$ denote the eigenvalues and eigenvectors for TT^*. Then

$$T = \sum_{k=1}^\infty \rho_i \tilde{g}_k \otimes_{\mathcal{H}(K_2)} \tilde{f}_k. \tag{27}$$

The k^{th} canonical correlation is ρ_k and $\{\tilde{f}_k, \tilde{g}_k\}_{k=1}^\infty$ are the canonical weight vectors in $\{\mathcal{H}(K_1), \mathcal{H}(K_2)\}$. These canonical weight vectors correspond to the canonical variables $\{\Psi_1(\tilde{f}_j), \Psi_Y(\tilde{g}_j)\}_{j=1}^\infty$ that represent the maximally correlated elements of $\{L^2_{X_1}, L^2_{X_2}\}$.

The relationship between T and R was established by Eubank and Hsing [15] and can be simply derived by substituting the expression for K_{12} given by (8) into (26). It follows that for any $\tilde{g} \in \mathcal{H}(K_2)$,

$$\begin{aligned}(T\tilde{g})(s) &= \sum_{j=1}^\infty \sum_{k=1}^\infty \gamma_{jk} \langle \phi_{2k}, \tilde{g}\rangle_{\mathcal{H}(K_2)} \phi_{1j}(s) \\ &= \sum_{j=1}^\infty \sum_{k=1}^\infty (\rho_{jk}\sqrt{\lambda_{1j}\lambda_{2k}}) \langle \phi_{2k}, \tilde{g}\rangle_{\mathcal{H}(K_2)} \phi_{1j}(s) \end{aligned} \tag{28}$$

with $\rho_{jk} = \frac{\gamma_{jk}}{\sqrt{\lambda_{1j}\lambda_{2k}}}$. Now, since $\{\phi_{ik}\}_{k=1}^\infty$ are CONSs for $\ker(S_i)^\perp$, it follows that $\{\tilde{\phi}_{ik} = \Gamma_i \phi_{ik} = \sqrt{\lambda_{ik}}\phi_{ik}\}_{k=1}^\infty$ are CONSs for $\mathcal{H}(K_i)$. As a result, the

operator T may be written as

$$\begin{aligned} T &= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\rho_{jk}\tilde{\phi}_{2k}\otimes_{\mathcal{H}(K_2)}\tilde{\phi}_{1j} \\ &= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\rho_{jk}[(\Gamma_2\phi_{2k})\otimes_{\mathcal{H}(K_2)}(\Gamma_1\phi_{1j})] \\ &= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\rho_{jk}[(\phi_{2k}\Gamma_2^{-1})\otimes_{\mathcal{H}_2}(\Gamma_1\phi_{1j})] \\ &= \Gamma_1\left[\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\rho_{jk}[\phi_{2k}\otimes_{\mathcal{H}_2}\phi_{1j}]\right]\Gamma_2^{-1} \\ &= \Gamma_1 R\Gamma_2^{-1} = \Gamma_1 S_1^{1/2\dagger}S_{12}S_2^{1/2\dagger}\Gamma_2^{-1}. \end{aligned} \tag{29}$$

Because Γ_2 is a bijection, $\Gamma_1^{-1}T\Gamma_2 : \mathcal{H}_2 \mapsto \mathcal{H}_1$ has the form

$$\Gamma_1^{-1}T\Gamma_2 = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\rho_{jk}\left(\phi_{2k}\otimes_{\mathcal{H}_2}\phi_{1j}\right)$$

and the domain is $\ker(S_2)^{\perp}$. By contrast, if we utilize the He et al. [17] method and restrict ourselves to the domain $F_2 = \mathrm{Im}(S_2^{1/2}) \subset \overline{\mathrm{Im}(S_2^{1/2})} = \ker(S_2)^{\perp}$ then, on this restricted subspace of $\ker(S_2)^{\perp}$,

$$R = S_1^{1/2\dagger}S_{12}S_2^{1/2\dagger} = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\rho_{jk}\left(\phi_{2k}\otimes_{\mathcal{H}_2}\phi_{1j}\right) = \Gamma_1^{-1}T\Gamma_2|_{F_2}.$$

Since Γ_1 and Γ_2 are unitary, T is unitarily equivalent to R and the two methods agree when both methods are well defined. The differences between the approaches can be briefly summarized by the fact that in the He et al. [17] approach the domain of R must be restricted to $F_2 = S_2^{1/2}\left(\ker(S_2)^{\perp}\right)$, which is a dense proper subset of $\ker(S_2)^{\perp}$ in the infinite dimensional case. By contrast, the domain of $\Gamma_1^{-1}T\Gamma_2$ in Eubank and Hsing [15] approach is all of $\ker(S_2)^{\perp}$, since the mapping Γ_2 is a unitary bijective mapping from $\ker(S_2)^{\perp} \mapsto \mathcal{H}(K_2)$. Therefore the Eubank and Hsing [15] approach is the more comprehensive definition while the He et al. [17] approach can have non-attainable solutions on the boundary $\left(\overline{\mathrm{Im}(S_2^{1/2})} \setminus \mathrm{Im}(S_2^{1/2})\right)$. This reveals the advantage of RKHS based formulation and we will therefore consider asymptotics associated with the regularized approximations to TT^* and T^*T rather than RR^* and R^*R in this paper.

4. Regularization

The need to employ some form of regularization in the functional data analysis setting is well established on both theoretical as well as computational

grounds by many authors. For example, it was perhaps Leurgans et al. [25] who first observed that the sample covariance operator of a stochastic process has a finite dimensional kernel (Riesz & Sz.-Nagy [32]), while acting on an infinite dimensional space. Cupidon et al. [6] then showed how most of the deficiencies of the population canonical correlation can be remedied if a regularized approximation to the inverses of the covariance operators are involved.

If $B \in \mathcal{K}(\mathcal{H}_1, \mathcal{H}_2)$ is arbitrary and we are given $g \in \mathcal{H}_2$, it often happens that we are asked to solve the equation $Bf = g$. If $\{\beta_n, \phi_n, \theta_n\}_{n=1}^{\mathrm{rank}(B^*B)}$ is the singular system for B so that

$$B = \sum_{n=1}^{\mathrm{rank}(B^*B)} \beta_n \phi_n \otimes_{\mathcal{H}_1} \theta_n$$

and $g \in \mathrm{Im}(B) \oplus \mathrm{Im}(B)^{\perp}$, then it is well known that a unique best approximate (least squares) solution f_* exists and is given by

$$f_* = B^{\dagger} g = \sum_{n=1}^{\mathrm{rank}(B^*B)} \frac{\langle y, \theta_n \rangle_{\mathcal{H}_2}}{\beta_n} \phi_n.$$

For a compact operator B, $Bf = g$ is often ill-posed (see e.g., Theorem 2.14 of Vogel [34]) and attempts to directly use $B^{\dagger}$ will result in numerically unstable algorithms. The standard approach to dealing with this problem is to replace $B^{\dagger}$ with a family of so called regularization operators $D(\alpha) : \mathcal{H}_2 \mapsto \mathcal{H}_1$ that are indexed by a regularization parameter, $\alpha \in (0, a) \subset \mathbb{R}$, with $a > 0$. The family $\{D(\alpha) : \alpha \in (0, a)\}$ approximates $B^{\dagger}$ in the sense of the following definition (see Vogel [34] p. 22-23).

Definition The family $\{D(\alpha) : \alpha \in (0, a)\}$ is a regularization scheme which converges to $B^{\dagger}$ if

(i) for each $\alpha \in (0, a)$, $D(\alpha)$ is a continuous operator and
(ii) given any $g \in \mathrm{Im}(B)$, for any sequence $\{g_n\} \subset \mathcal{H}_2$ which converges to g, one can pick a sequence $\{\alpha_n\} \subset (0, a)$ such that

$$[D(\alpha_n)](g_n) \to B^{\dagger} g \quad \text{as } n \to \infty.$$

Of particular interest are linear regularization schemes which have singular value representations as

$$D(\alpha) = \sum_{n=1}^{\mathrm{rank}(B^*B)} \frac{w_\alpha(\beta_n^2)}{\beta_n} \theta_n \otimes_{\mathcal{H}_2} \phi_n$$

where $w_\alpha(\beta_n^2)$ is a real valued function of the squared singular values and α is such that $w_\alpha(\beta_n^2) \to 1$ as $\alpha \to 0$. The function $w_\alpha(\beta_n^2)$ is called the filter function (see Engl et al. [14]). Two of the most popular examples for filters are

$$w_\alpha(\beta_n^2) = \frac{|\beta_n|}{|\beta_n| + \alpha}, \quad \text{for } \alpha \in (0, \infty) \text{ and } n = 1, \dots, \mathrm{rank}(B^*B) \tag{30}$$

and

$$w_\alpha(\beta_n^2) = \begin{cases} 1 & \text{if } \beta_n^2 > \alpha \\ 0 & \text{if } \beta_n^2 \le \alpha \end{cases} \quad \text{for } \alpha \in (0, \|B\|] \text{ and } n = 1, \ldots, \operatorname{rank}(B^*B). \tag{31}$$

Equation (30) is referred to as the Tikhinov filter function and (31) is referred to as the truncated singular value decomposition (TSVD) filter function. In the case of TSVD regularization, the parameter α in (31) determines the cut-off or threshold level for the TSVD regularization and produces

$$D(\alpha) = \sum_{\beta_n^2 > \alpha} \beta_n^{-1} \theta_n \otimes_{\mathcal{H}_2} \phi_n$$

which is a finite rank operator whenever $\alpha > 0$. This paper will focus on asymptotics associated with Tikhinov and TSVD regularization schemes. In developments which follow a Hilbert-Schmidt operator $B : \mathcal{H}_1 \mapsto \mathcal{H}_2$ will often be expressed in the form

$$B = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} c_{jk} \phi_j \otimes_{\mathcal{H}_1} \theta_k \tag{32}$$

with $\{\phi_j\}$, $\{\theta_k\}$ CONSs for $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. As $B \in \mathcal{K}_{HS}(\mathcal{H}_1, \mathcal{H}_2)$ the coefficients $c_{jk} \in \mathbb{R}$ will satisfy

$$\|B\|_{HS}^2 = \sum_{j=1}^{\infty} \langle B\phi_j, B\phi_j \rangle_{\mathcal{H}_1} = \sum_{j=1}^{\infty} \| \sum_{k=1}^{\infty} c_{jk}^2 \theta_k \|_{\mathcal{H}_2} = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} c_{jk}^2 < \infty.$$

The regularized operator $B(\alpha) : \mathcal{H}_1 \mapsto \mathcal{H}_2$ will also be a Hilbert-Schmidt operator and have the form

$$B(\alpha) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} c_{jk}(\alpha) \phi_j \otimes_{\mathcal{H}_1} \theta_k \tag{33}$$

with $\|B(\alpha)\|_{HS}^2 = \sum_{i,j=1}^{\infty} c_{ij}^2(\alpha) < \infty$ and the property $c_{jk}(\alpha) \to c_{jk}$ as $\alpha \downarrow 0$. As a result, the following theorem holds.

Theorem 4.1. *Suppose $B, B(\alpha) \in \mathcal{K}_{HS}(\mathcal{H}_1, \mathcal{H}_2)$ are of the forms (32) and (33), respectively. If $c_{jk}(\alpha) \to c_{jk}$ as $\alpha \to 0$, then $\|B(\alpha) - B\| \to 0$.*

Proof: Let $A(\alpha) = B(\alpha) - B$ and $a_{jk}(\alpha) = (c_{jk}(\alpha) - c_{jk})$ so that

$$A(\alpha) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} a_{jk}(\alpha) \phi_j \otimes_{\mathcal{H}_1} \theta_k.$$

Since $A(\alpha)$ is Hilbert-Schmidt, $\|A(\alpha)\|_{HS}^2 = \sum_{j,k=1}^{\infty} a_{jk}^2(\alpha) < \infty$ for all permissable values of the regularization parameter α. Consequently,

$$\begin{aligned} \lim_{\alpha \to 0} \|A(\alpha)\|_{HS}^2 &= \lim_{\alpha \to 0} \sum_{j,k=1}^{\infty} a_{jk}^2(\alpha) \\ &= \sum_{j,k=1}^{\infty} \lim_{\alpha \to 0} a_{jk}^2(\alpha) = 0. \end{aligned} \tag{34}$$

The exchange in the order of limits and the sum in (34) is permissable by the Lebesgue dominated convergence theorem since the summands satisfy $a_{jk}^2(\alpha) \leq 2(c_{jk}^2(\alpha) + c_{jk}^2)$ and $\sum_{j,k=1}^{\infty} a_{jk}^2(\alpha) < \infty$. Now, since $\|B(\alpha) - B\|^2 \leq \|B(\alpha) - B\|_{HS}^2 = \|A(\alpha)\|_{HS}^2 \to 0$ as $\alpha \downarrow 0$, it follows that $B(\alpha)$ converges to B in operator norm. ♢

5. Tikhinov Regularized Canonical Correlation

In the Tikhinov regularized approach to canonical correlation we replace the operators $\{S_i^{1/2\dagger}\}_{i=1}^2$ with $\{(S_i + \alpha I)^{-1/2}\}_{i=1}^2$ and then let

$$R(\alpha) \equiv (S_1 + \alpha I)^{-1/2} S_{12} (S_2 + \alpha I)^{-1/2} \tag{35}$$

approximate the cross-correlation operator R for $\alpha \in (0, a)$. Since the operators $\{(S_1+\alpha I)^{-1/2}, (S_2+\alpha I)^{-1/2}\}$ are bounded and S_{12} is Hilbert-Schmidt, it follows that $R(\alpha)$ is Hilbert-Schmidt. Now if we define $T(\alpha) : \mathcal{H}(K_2) \mapsto \mathcal{H}(K_1)$ by

$$T(\alpha) \equiv \Gamma_1 R(\alpha) \Gamma_2^{-1} = \Gamma_1 (S_1 + \alpha I)^{-1/2} S_{12} (S_2 + \alpha I)^{-1/2} \Gamma_2^{-1} \tag{36}$$

then as the regularization parameter $\alpha \downarrow 0$,

$$\begin{aligned} R(\alpha) &= (S_1 + \alpha I_1)^{-1/2} S_{12} (S_2 + \alpha I_2)^{-1/2} \\ &= \sum_{j=1}^{\mathrm{rank}(S_1)} \sum_{k=1}^{\mathrm{rank}(S_2)} \frac{\gamma_{jk}}{\sqrt{(\lambda_{1j} + \alpha)(\lambda_{2k} + \alpha)}} \phi_{2k} \otimes_{L^2(E_2)} \phi_{1j} \\ &\to \sum_{j=1}^{\mathrm{rank}(S_1)} \sum_{k=1}^{\mathrm{rank}(S_2)} \frac{\gamma_{jk}}{\sqrt{\lambda_{1j}\lambda_{2k}}} \phi_{2k} \otimes_{L^2(E_2)} \phi_{1j} \\ &= \Gamma_1^{-1} T \Gamma_2. \end{aligned} \tag{37}$$

Therefore, by Therem 4.1, $R(\alpha)$ converges in operator norm to the operator $\Gamma_1^{-1} T \Gamma_1$ as $\alpha \downarrow 0$ and the continuity of Γ_1 and Γ_2 ensures that

$$T(\alpha) = \Gamma_1 R(\alpha) \Gamma_2^{-1} \to T$$

as $\alpha \downarrow 0$, with convergence in terms of operator norm.

We will now show that the regularized canonical correlations along with the regularized canonical variables converge to the canonical correlation and variables defined from the Eubank and Hsing [15] methodology as $\alpha \downarrow 0$. In this regard, suppose that $\{\rho_k(\alpha), f_k(\alpha), g_k(\alpha)\}_{k=1}^{\infty}$ is the singular system for $R(\alpha)$ such that

$$R(\alpha) = \sum_{k=1}^{\infty} \rho_k(\alpha) \left[g_k(\alpha) \otimes_{\mathcal{H}_2} f_k(\alpha) \right].$$

Then,

$$
\begin{aligned}
T(\alpha) = \Gamma_1 R(\alpha)\Gamma_2^{-1} &= \sum_{k=1}^{\infty} \rho_k(\alpha) \left[(g_k(\alpha)\Gamma_2^*) \otimes_{\mathcal{H}(K_2)} (\Gamma_1 f_k(\alpha)) \right] \\
&= \sum_{k=1}^{\infty} \rho_k(\alpha) \left[\tilde{g}_k(\alpha) \otimes_{\mathcal{H}(K_2)} \tilde{f}_k(\alpha) \right]
\end{aligned}
$$

where $\tilde{f}_k(\alpha) = \Gamma_1 f_k(\alpha) \in \mathcal{H}(K_1)$ and $\tilde{g}_k(\alpha) = \Gamma_2 g_k(\alpha) \in \mathcal{H}(K_2)$. Now by (16) the canonical weight functions may be written as

$$
\tilde{f}_k(\alpha) = \sum_{j=1}^{\infty} \lambda_{1j} f_{kj}(\alpha)\phi_{1j} \quad \text{and} \quad \tilde{g}_k(\alpha) = \sum_{j=1}^{\infty} \lambda_{2j} g_{kj}(\alpha)\phi_{2j}
$$

with

$$
f_{kj}(\alpha) = \langle \tilde{f}_k(\alpha), \phi_{1j} \rangle_{\mathcal{H}_1} \quad \text{and} \quad g_{kj}(\alpha) = \langle \tilde{g}_k(\alpha), \phi_{2j} \rangle_{\mathcal{H}_2}.
$$

Utilizing Theorem (2.1) the corresponding regularized canonical variables in $L^2_{X_1}$ and $L^2_{X_2}$ are then

$$
\begin{aligned}
U_k(\alpha) &= \Psi_1(\tilde{f}_k(\alpha)) = \sum_{j=1}^{\operatorname{rank}(S_1)} f_{kj}(\alpha) \langle X_1, \phi_{1j} \rangle_{\mathcal{H}_1} \text{ and} \\
V_k(\alpha) &= \Psi_2(\tilde{g}_k(\alpha)) = \sum_{j=1}^{\operatorname{rank}(S_2)} g_{kj}(\alpha) \langle X_2, \phi_{2j} \rangle_{\mathcal{H}_2}.
\end{aligned}
$$

The continuity of the congruence mappings Ψ_1 and Ψ_2 ensures the convergence of the regularized canonical variables $U_k(\alpha) = \Psi_1(\tilde{f}_k(\alpha))$ and $V_k(\alpha) = \Psi_2(\tilde{g}_k(\alpha))$ to the true canonical variables, provided that the regularized canonical weight functions $\tilde{f}_k(\alpha) \in \mathcal{H}(K_1)$ and $\tilde{g}_k(\alpha) \in \mathcal{H}(K_2)$ converge to the true canonical weight functions $\tilde{f}_k$ and $\tilde{g}_k$ as the regularization parameter tends to zero. Thus, our tasks are to establish convergence of $\rho_k^2(\alpha)$ to ρ_k^2 and of the the regularized RKHS functions $\{\tilde{f}_k(\alpha), \tilde{g}_k(\alpha)\}$ to $\{\tilde{f}_k, \tilde{g}_k\}$ for all $k \geq 1$. Concerning the convergence of the eigenvalues $\rho_k^2(\alpha)$ and the corresponding eigenprojection operators we have the following result.

Theorem 5.1. *Let $\{\rho_k^2(\alpha), P_k(\alpha)\}$ and $\{\rho_k^2, P_k\}$ denote the eigenvalues and corresponding eigenprojection operators for $T(\alpha)T^*(\alpha)$ and TT^*. The following limits hold as $\alpha \downarrow 0$*

$$
0 \leq \rho_k^2(\alpha) \uparrow \rho_k^2 \leq 1 \text{ as } \alpha \downarrow 0 \text{ for all } k \geq 1 \text{ and} \tag{38}
$$

$$
\|P_k(\alpha) - P_k\| \leq \|P_k(\alpha) - P_k\|_{HS} \to 0. \tag{39}
$$

Proof: First note that $\rho_k^2(\alpha) < \rho_k^2 < 1$ since $\|T(\alpha)\|^2 < \|T\|^2 \leq 1$. (see Proposition A.3 of Eubank and Hsing [15]). Now to see that (38) holds, fix $k \geq 1$. Since $\rho_k^2(\alpha)$ and ρ_k^2 are the k^{th} eigenvalues for $T(\alpha)T^*(\alpha)$ and TT^*

$$\left(\rho_k^2 - \rho_k^2(\alpha)\right) = |\rho_k^2 - \rho_k^2(\alpha)| \leq \|TT^* - T(\alpha)T^*(\alpha)\| \downarrow 0$$

as $\alpha \downarrow 0$. In order to show that $P_k(\alpha) \to P_k$ in operator norm, let $\Gamma_{r,k}$ be a circle centered at ρ_k^2 with radius r chosen so that $\Gamma_{r,k}$ encloses ρ_k^2 and no other eigenvalues of TT^*. Suppose that $\{R(\alpha, z), R(z)\}$ are the resolvents of $\{T(\alpha)T^*(\alpha), TT^*\}$, respectively. Since $\|T(\alpha)T^*(\alpha) - TT^*\| \to 0$ as $\alpha \downarrow 0$, it follows from Theorem 10.1 in the appendix that there exists $\alpha_0 > 0$ such that whenever $0 < \alpha < \alpha_0$, $\Gamma_{r,k}$ encloses $\rho_k^2(\alpha)$ and no other eigenvalues of $T(\alpha)T^*(\alpha)$. Furthermore, for any $\epsilon > 0$ we may take α_0 to be sufficiently small to ensure that $\|T(\alpha)T^*(\alpha) - TT^*\| < \epsilon$. Relation (92) from the appendix then has the consequence that

$$\|P_k(\alpha) - P_k\|_{HS} \leq r \sup_{z \in \Gamma_{r,k}} \left\{ \frac{\|T(\alpha)T^*(\alpha) - TT^*\|_{HS} \|R(z)\|_{HS}^2}{1 - \|T(\alpha)T^*(\alpha) - TT^*\|_{HS} \|R(z)\|_{HS}} \right\}.$$

Thus, if $M(r,k) \equiv \sup_{z \in \Gamma_{r,k}} \|R(z)\|_{HS}$ and $\epsilon > 0$ are chosen so that $\epsilon < \frac{1}{2M(r,k)}$, $\|P_k(\alpha) - P_k\|_{HS} \leq 2rM^2(r,k)\epsilon$ and hence $\|P_k(\alpha) - P_k\| < \|P_k(\alpha) - P_k\|_{HS} \to 0$ as $\alpha \downarrow 0$. $\diamondsuit$

It remains to show that for $k \geq 1$, $\tilde{f}_k(\alpha) = \Gamma_1(f_k(\alpha))$ and $\tilde{g}_k(\alpha) = \Gamma_2(g(\alpha))$ approach $\tilde{f}_k \in \mathcal{H}(K_1)$ and $\tilde{g}_k \in \mathcal{H}(K_2)$ from the singular system $\{\rho_k, \tilde{f}_k, \tilde{g}_k\}$ of T. We note, however, that eigenvectors associated with any operator are not defined uniquely. For example, if θ is an eigenvector for an arbitrary self-adjoint operator A, then $-\theta$ is also an eigenvector. In order to properly establish what we mean by convergence assume, WLOG, that for all $\alpha > 0$, $\tilde{f}_k(\alpha)$ be chosen so that $\langle \tilde{f}_k(\alpha), \tilde{f}_k \rangle_{\mathcal{H}(K_1)} \geq 0$, with a similar convention applied to $\tilde{g}_k(\alpha)$. The theorem below concerns convergence in the case that the eigenspaces associated with $\{\tilde{f}_k, \tilde{g}_k\}$ are 1–dimensional. Subsequently, we will discuss the higher dimensional case.

Theorem 5.2. *Assume that the eigenspaces associated with the eigenvectors $\tilde{f}_k$ and $\tilde{g}_k$ are one dimensional with $k \in \mathbb{Z}$. Then, as $\alpha \downarrow 0$*

$$\|\tilde{f}_k(\alpha) - \tilde{f}_k\|_{\mathcal{H}(K_1)} \to 0 \quad \textit{and} \quad \|\tilde{g}_k(\alpha) - \tilde{g}_k\|_{\mathcal{H}(K_2)} \to 0.$$

Proof: For fixed $k \in \mathbb{Z}$, it suffices to show that $\|\tilde{f}_k(\alpha) - \tilde{f}_k\|_{\mathcal{H}(K_1)} \to 0$. Since the eigenspaces are one-dimensional it follows that $P_k(\alpha) = \left[\tilde{f}_k(\alpha) \otimes_{\mathcal{H}(K_1)} \tilde{f}_k(\alpha)\right]$

and $P_k = \left[\tilde{f}_k \otimes_{\mathcal{H}(K_1)} \tilde{f}_k\right]$. Now, notice that

$$\begin{aligned}
\|P_k(\alpha) - P_k\|_{HS} &= \langle P_k(\alpha) - P_k, P_k(\alpha) - P_k\rangle_{HS} \\
&= 2 - 2\langle P_k(\alpha), P_k\rangle_{HS} \\
&= 2 - 2\langle \tilde{f}_k(\alpha) \otimes_{\mathcal{H}(K_1)} \tilde{f}_k(\alpha), \tilde{f}_k \otimes_{\mathcal{H}(K_1)} \tilde{f}_k\rangle_{HS} \\
&= 2 - 2\langle \tilde{f}_k(\alpha), \tilde{f}_k\rangle^2_{\mathcal{H}(K_1)} \\
&= 2(1 - \langle \tilde{f}_k(\alpha), \tilde{f}_k\rangle_{\mathcal{H}(K_1)})(1 + \langle \tilde{f}_k(\alpha), \tilde{f}_k\rangle_{\mathcal{H}(K_1)}) \\
&= \|\tilde{f}_k(\alpha) - \tilde{f}_k\|^2_{\mathcal{H}(K_1)}(1 + \langle \tilde{f}_k(\alpha), \tilde{f}_k\rangle_{\mathcal{H}(K_1)}). \qquad (40)
\end{aligned}$$

Furthermore, as $\langle \tilde{f}_k(\alpha), \tilde{f}_k\rangle_{\mathcal{H}(K_1)} \geq 0 \longrightarrow (1 + \langle \tilde{f}_k(\alpha), \tilde{f}_k\rangle_{\mathcal{H}(K_1)}) \geq 1$, hence

$$\|\tilde{f}_k(\alpha) - \tilde{f}_k\|^2_{\mathcal{H}(K_1)} \leq \|P_k(\alpha) - P_k\|_{HS}.$$

Since $\|P_k(\alpha) - P_k\|_{HS} \to 0$ as $\alpha \downarrow 0$, it follows that $\|\tilde{f}_k(\alpha) - \tilde{f}_k\|^2_{\mathcal{H}(K_1)} \to 0$. $\diamondsuit$

It should be noted that if $\langle \tilde{f}_k(\alpha), \tilde{f}_k\rangle_{\mathcal{H}(K_1)} \leq 0$ instead, then $(1 - \langle \tilde{f}_k(\alpha), \tilde{f}_k\rangle_{\mathcal{H}(K_1)}) \geq 1$ and from (40) we would have

$$\begin{aligned}
\|P_k(\alpha) - P_k\|_{HS} &= 2(1 + \langle \tilde{f}_k(\alpha), -\tilde{f}_k\rangle_{\mathcal{H}(K_1)})(1 - \langle \tilde{f}_k(\alpha), -\tilde{f}_k\rangle_{\mathcal{H}(K_1)}) \\
&= (1 - \langle \tilde{f}_k(\alpha), \tilde{f}_k\rangle_{\mathcal{H}(K_1)})\|\tilde{f}_k(\alpha) - (-\tilde{f}_k)\|^2_{\mathcal{H}(K_1)}.
\end{aligned}$$

Hence,

$$\|\tilde{f}_k(\alpha) - (-\tilde{f}_k)\|^2_{\mathcal{H}(K_1)} \leq \|P_k(\alpha) - P_k\|_{HS} \to 0 \quad \text{as} \quad \alpha \downarrow 0$$

and $\tilde{f}_k(\alpha)$ would converge to $(-\tilde{f}_k)$ instead.

When the eigenspaces have dimension larger than 1, it is possible to find infinitely many eigenspace invariant rotations $\Theta \in \mathcal{B}(\mathcal{H}(K_1))$ so that $\tilde{f}'_k = \Theta\tilde{f}_k$ is still an eigenvector of TT^* with eigenvalue ρ_k^2, yet $\|\tilde{f}_k(\alpha) - \tilde{f}'_k\| \nrightarrow 0$ as $\alpha \downarrow 0$ (see Kato [22] p. 98-100).

Theorems 5.1 and 5.2 ensure that when T is simple, the singular system $\{\rho_k(\alpha), \tilde{f}_k(\alpha), \tilde{g}_k(\alpha)\}$ of $T(\alpha)$ converges to the singular system $\{\rho_k, \tilde{f}_k, \tilde{g}_k\}$ of T as the regularization parameter $\alpha \downarrow 0$. This is a positive development provided the singular value decomposition of $T(\alpha)$ can be estimated. However, the singular value decomposition of $T(\alpha)$ entails the eigenvalue-eigenvector decomposition of e.g., the operator

$$\begin{aligned}
\mathcal{T}_1(\alpha) &\equiv T(\alpha)T^*(\alpha) = \Gamma_1 R(\alpha)R^*(\alpha)\Gamma_1^{-1} \\
&= \Gamma_1(S_1 + \alpha I)^{-1/2}S_{12}(S_2 + \alpha I)^{-1}S_{21}(S_1 + \alpha I)^{-1/2}\Gamma_1^{-1}. \qquad (41)
\end{aligned}$$

Since, Γ_1 is unknown in (41) we might estimate it using

$$\hat{\Gamma}_{1n}(m) = \sum_{i=1}^{m} \sqrt{\hat{\lambda}_{1in}}\hat{P}_{1in}$$

with $\{\hat{\lambda}_{1in}, \hat{P}_{1in}\}$, the estimated eigenvalues and corresponding eigenprojection operators for $\hat{S}_{1n}$ and m some integer. This raises the question of how to select m and, for large m, $\hat{\Gamma}_{1n}(m)$ is approximately $\hat{S}_{1n}^{1/2}$ whose compact nature is what prompted us to regularize from the beginning.

Since we are already utilizing Tikhinov regularization, a possible remedy for our problem is to replace Γ_1 with $(S_1 + \alpha I)^{1/2}$. This produces the operator

$$\mathcal{S}_1(\alpha) \equiv S_{12}(S_2 + \alpha I)^{-1} S_{21}(S_1 + \alpha I)^{-1} \tag{42}$$

whose domain is $\ker(S_1)^{\perp}$ rather than $\mathcal{H}(K_1)$. One advantage of $\mathcal{S}_1(\alpha)$ is that $\mathrm{Im}(\mathcal{S}_1(\alpha)) \subseteq \mathrm{Im}(S_1) \subseteq F_1$ and hence the eigenfunctions of $\mathcal{S}_1(\alpha)$ satisfy the Picard criteria. To see this, note by the infinite dimensional extension of the result from Khatri [20] we have that $S_1 S_1^{\dagger} S_{12} = S_{12}$ and hence $\mathcal{S}_1(\alpha) = S_1 S_1^{\dagger} \mathcal{S}_1(\alpha)$ (see King [20]). Note that the operator $\mathcal{S}_1(\alpha)$ is self-adjoint since

$$\begin{aligned} \mathcal{S}_1(\alpha) &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{\gamma_{jk}^2}{(\lambda_{1j} + \alpha)(\lambda_{2k} + \alpha)} \left[\phi_{1j} \otimes_{\mathcal{H}_1} \phi_{1k}\right] \\ &= (S_1 + \alpha I)^{-1} S_{12} (S_2 + \alpha I)^{-1} S_{21} = \mathcal{S}_1^*(\alpha). \end{aligned}$$

Furthermore, as S_{12} is a factor in $\mathcal{S}_1(\alpha)$, the operator is Hilbert-Schmidt and hence admits an eigenvalue-eigenvector decomposition

$$\mathcal{S}_1(\alpha) = \sum_{j=1}^{\infty} \rho_j^2(\alpha) \left[f_j(\alpha) \otimes_{\mathcal{H}_1} f_j(\alpha)\right].$$

Now the question becomes how can an operator whose domain and range are subsets of $\mathcal{H}_1$, approximate an operator whose domain and range are subsets of $\mathcal{H}(K_1)$. The answer to this question was fundamentally answered by Nasheed and Wahba [27] when it was proved that the collection of functions in $\mathcal{H}(K_1)$ is the same as $\overline{\mathrm{Im}(S_1)^{1/2}}$, except with alternate norm and inner product. As the collection of eigenfunctions $\{f_j(\alpha)\}_{j=1}^{\infty}$ reside in $\mathrm{Im}(S_1)^{1/2}$, they also have "dual citizenship" in $\mathcal{H}(K_1)$. We may therefore regard the eigenfunction sequence $\{f_j(\alpha)\}$ as residing in $\mathcal{H}(K_1)$, provided that we norm the eigenfunctions correctly. If we treat the eigenfuctions $f_j(\alpha)$ as citizens of $\mathcal{H}(K_1)$, for notational consistency we will denote them by $\tilde{f}_j(\alpha)$ with $\{f_j(\alpha) = \tilde{f}_j(\alpha)\}_{j=1}^{\mathrm{rank}(\mathcal{S}_1(\alpha))}$. There are therefore two possible views one may adopt concerning the operator $\mathcal{S}_1(\alpha)$:

- (i) In the first view of $\mathcal{S}_1(\alpha)$, we treat the operator as a self-adjoint mapping in $\ker(S_1)^{\perp} \subseteq L^2(E_1)$.
- (ii) In the second viewpoint, the operator is treated as a self-adjoint mapping on $\mathcal{H}(K_1)$ with $\mathcal{S}_1(\alpha)$ regarded as "two perturbations" distant from the the operator TT^*, which is its ultimate intended target of approximation.

When the second viewpoint for $\mathcal{S}_1(\alpha)$ is adopted, the operator $\mathcal{S}_1(\alpha)$ is representable by the $\mathcal{H}(K_1)$ based operator

$$\mathcal{S}_1(\alpha) = \Gamma_1 S_1^{1/2\dagger} S_{12}(S_2 + \alpha I)^{-1} S_{21}(S_1 + \alpha I)^{-1} S_1^{1/2} \Gamma_1^{-1}.$$

We then see that as $\alpha \downarrow 0$,

$$\begin{aligned}\Gamma_1^{-1}\mathcal{S}_1(\alpha)\Gamma_1 &= S_1^{1/2\dagger} S_{12}(S_2+\alpha I)^{-1}S_{21}(S_1+\alpha I)^{-1}S_1^{1/2} \\ &= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\frac{\gamma_{jk}^2}{(\lambda_{1j}+\alpha)(\lambda_{2k}+\alpha)}\left[\phi_{1j}\otimes_{\mathcal{H}_1}\phi_{1k}\right] \\ &\to \sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\frac{\gamma_{jk}^2}{(\lambda_{1j})(\lambda_{2k})}\left[\phi_{1j}\otimes_{\mathcal{H}_1}\phi_{1k}\right] \\ &= \Gamma_1^{-1}TT^*\Gamma_1.\end{aligned}$$

Therefore, since $\mathcal{S}_1(\alpha)$ is Hilbert-Schmidt, Theorem 4.1 ensures that as the regularization parameter $\alpha \downarrow 0$,

$$\|\Gamma_1^{-1}\mathcal{S}_1(\alpha)\Gamma_1 - \Gamma_1^{-1}TT^*\Gamma_1\|_{\mathcal{H}_1} = \|\mathcal{S}_1(\alpha) - TT^*\|_{\mathcal{H}(K_1)} \to 0.$$

6. Asymptotic Properties for Tikhinov Regularization

In this section we will consider the asymptotics associated with the sample estimators of the operators $\mathcal{S}_1(\alpha)$. The asymptotics associated with the operator $\mathcal{S}_1(\alpha)$ rely heavily on perturbation theory concepts discussed in Dauxois et al. [9] as well as delta method theory for random operators discussed in Cupidon et al. [7].

To begin, we suppose that a random sample $X_1, X_2, \ldots X_n$ of independent, identically distributed copies of $X \in L^2(E)$ are observed. The sample estimator associated with the covariance operator of X is given by

$$\hat{S}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}) \otimes_{\mathcal{H}} (X_i - \bar{X}) \tag{43}$$

and the continuous mapping theorem along with the law of large numbers ensures that $\hat{S}_{in} = \Upsilon_i\hat{S}_n\Upsilon_i \xrightarrow{p} S_i$ for $i = 1, 2$ and $\hat{S}_{12n} = \Upsilon_1\hat{S}_n\Upsilon_2 = \hat{S}_{21n}^* \xrightarrow{p} S_{12}$ as $n \to \infty$. For Tikhinov regularization we will have need of the function $\varphi_\alpha(z) \equiv (z+\alpha)^{-1}$, which is analytic for all points in the complex plane, except for a pole at $z = -\alpha$. Consequently, the disk $D = \{z \in \mathbb{C} | \min_{0\le x\le \|S\|} |z - x| < \frac{\alpha}{2}\}$ contains the spectra of S and the function φ_α is analytic on D. It follows by the continuous mapping theorem that $\varphi_\alpha(\hat{S}_{in}) \xrightarrow{p} \varphi_\alpha(S_i)$ as $n \to \infty$ for $i = 1, 2$.

As a consequence of the continuous mapping theorem and the central limit theorem for Hilbert space operators (see Dauxois et al. [9]) we have that

$$\sqrt{n}(\Upsilon_i\hat{S}_n\Upsilon_j - \Upsilon_i S\Upsilon_j) = \sqrt{n}(\hat{S}_{ijn} - S_{ij}) \xrightarrow{d} \Upsilon_i\mathcal{N}\Upsilon_j = \mathcal{N}_{ij} \tag{44}$$

where, for $i, j = 1, 2$, $\mathcal{N}_{ij} \in \mathcal{K}_{HS}(\mathcal{H}_i, \mathcal{H}_j)$ is a Gaussian random operator that has mean zero and variance

$$\Sigma_{ij} \equiv \mathrm{E}\{(X_i \otimes_{\mathcal{H}_i} X_i - S_i) \otimes_{HS} (X_j \otimes_{\mathcal{H}_j} X_j - S_j)\} \tag{45}$$

and $\mathcal{N}_{ii} \equiv \mathcal{N}_i$. Furthermore, by the delta method result from Cupidon et al. [7] it follows that for $i = 1, 2$

$$\sqrt{n}\left\{\varphi_\alpha(\hat{S}_{in}) - \varphi_\alpha(S_i)\right\} \xrightarrow{d} \varphi'_\alpha(\mathcal{N}_i) \tag{46}$$

where the limit in $\mathcal{K}_{HS}(\mathcal{H}_i)$ has zero mean and is distributed as

$$\begin{aligned} \varphi'_\alpha(\mathcal{N}_i) = & - \sum_{k=1}^{\infty}(\lambda_{ik} + \alpha)^{-2} P_{ik}\mathcal{N}_i P_{ik} \\ & + \sum_{j \neq k} \frac{1}{(\lambda_{ik} + \alpha)(\lambda_{ij} + \alpha)} P_{ij}\mathcal{N}_i P_{ik} \end{aligned} \tag{47}$$

with $\{\lambda_{ik}, P_{ik}\}_{k=1}^{\infty}$, the eigenvalues and eigenprojection operators corresponding to S_i, $i = 1, 2$ (see Appendix).

The sample version of the operator $\mathcal{S}_1(\alpha)$ is then defined by

$$\hat{\mathcal{S}}_{1n}(\alpha) \equiv \hat{S}_{12n}(\hat{S}_{2n} + \alpha I)^{-1}\hat{S}_{21n}(\hat{S}_{1n} + \alpha I)^{-1}. \tag{48}$$

The asymptotic analysis of $\sqrt{n}(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha))$ follows from a product rule application of the delta method similar to that in Cupidon, et al. [7]. In this regard we introduce the following Gaussian elements in the set $\mathcal{K}_{HS}(\mathcal{H}(K_1))$ of Hilbert-Schmidt operators on $\mathcal{H}(K_1)$

$$\begin{aligned} \mathcal{G}_{11}(\alpha) & \equiv \mathcal{N}_{12}\varphi_\alpha(S_2)S_{21}\varphi_\alpha(S_1), \\ \mathcal{G}_{12}(\alpha) & \equiv S_{12}\varphi'_\alpha(\mathcal{N}_2)S_{21}\varphi_\alpha(S_1), \\ \mathcal{G}_{13}(\alpha) & \equiv S_{12}\varphi_\alpha(S_2)\mathcal{N}_{21}\varphi_\alpha(S_1), \\ \mathcal{G}_{14}(\alpha) & \equiv S_{12}\varphi_\alpha(S_2)S_{21}\varphi'_\alpha(\mathcal{N}_1), \\ \mathcal{G}_1(\alpha) & \equiv \sum_{k=1}^{4}\mathcal{G}_{1k}(\alpha). \end{aligned} \tag{49}$$

Corollary 6.1. *If* $\mathrm{E}\|X\|_{L^2(E)}^4 < \infty$*, then as* $n \to \infty$

$$\sqrt{n}(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha)) \xrightarrow{d} \mathcal{G}_1(\alpha). \tag{50}$$

Proof: Define

$$\begin{aligned} \hat{\mathcal{A}}_{11}(\alpha) & \equiv \left[\hat{S}_{12n} - S_{12}\right]\varphi_\alpha(\hat{S}_{2n})\hat{S}_{21n}\varphi_\alpha(\hat{S}_{1n}), \\ \hat{\mathcal{A}}_{12}(\alpha) & \equiv S_{12}\left[\varphi_\alpha(\hat{S}_{2n}) - \varphi_\alpha(S_2)\right]\hat{S}_{21n}\varphi_\alpha(\hat{S}_{1n}), \\ \hat{\mathcal{A}}_{13}(\alpha) & \equiv S_{12}\varphi_\alpha(S_2)\left[\hat{S}_{21n} - S_{21}\right]\varphi_\alpha(\hat{S}_{1n}), \\ \hat{\mathcal{A}}_{14}(\alpha) & \equiv S_{12}\varphi_\alpha(S_2)S_{21}\left[\varphi_\alpha(\hat{S}_{1n}) - \varphi_\alpha(S_1)\right]. \end{aligned}$$

Notice that the difference $\sqrt{n}(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha))$ can be expanded so that

$$\sqrt{n}(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha)) = \sqrt{n}\left[\sum_{j=1}^{4} \hat{\mathcal{A}}_{1j}(\alpha)\right].$$

The application of (44), (46) and Slutsky's Theorem then ensure that

$$\sqrt{n}\left[\sum_{j=1}^{4} \hat{\mathcal{A}}_{1j}(\alpha)\right] \xrightarrow{d} \mathcal{G}_1(\alpha)$$

since, for example, the term $\hat{\mathcal{A}}_{11}(\alpha)$ consists of the factor $\sqrt{n}\left[\hat{S}_{12n} - S_{12}\right] \xrightarrow{d} \mathcal{N}_{12}$ right-multiplied by the factor

$$\varphi_\alpha(\hat{S}_{2n})\hat{S}_{21n}\varphi_\alpha(\hat{S}_{1n}) \xrightarrow{p} \varphi_\alpha(S_2)S_{21}\varphi_\alpha(S_1).$$

$\diamondsuit$

As a result of Corollary 6.1, we see that $\hat{\mathcal{S}}_{1n}(\alpha)$ is a consistent estimator of $\mathcal{S}_1(\alpha)$ as

$$\|\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha)\| = \mathcal{O}_P(n^{-1/2}) \xrightarrow{p} 0. \tag{51}$$

However, note that as long as the regularization parameter $\alpha > 0$, $\|\hat{\mathcal{S}}_{1n}(\alpha) - TT^*\| \overset{p}{\nrightarrow} 0$. In fact, by the triangle inequality we have

$$\|\hat{\mathcal{S}}_{1n}(\alpha) - TT^*\| \leq \|\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha)\| + \|\mathcal{S}_1(\alpha) - TT^*\|. \tag{52}$$

The first term on the right-hand side of (52) can be viewed as a random error that originates from using a sample estimator of $\mathcal{S}_1(\alpha)$. This term tends to zero as $n \to \infty$ by (51). On the other hand, the second term on the right hand side of (52) is a deterministic error that arises from the regularized approximation of TT^*. This latter term will only become negligible if $\alpha \downarrow 0$.

Since the limiting distribution for $\sqrt{n}(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha))$ has been established, we may establish the limiting distributions associated with sample estimators for the k^{th} regularized canonical correlation and associated projection operator and weight functions. The quantities of interest are $\sqrt{n}\left\{\hat{\rho}_{kn}(\alpha) - \rho_k(\alpha)\right\}$, $\sqrt{n}\left\{\hat{\tilde{P}}_{1kn}(\alpha) - \tilde{P}_{1k}(\alpha)\right\}$ and $\sqrt{n}\left\{\hat{\tilde{f}}_{kn}(\alpha) - \tilde{f}_k(\alpha)\right\}$ where $\{\rho_k(\alpha), \tilde{P}_{1k}(\alpha), \tilde{f}_k(\alpha)\}$ denote the eigenvalues, eigenprojections and eigenvectors for $\mathcal{S}_1(\alpha)$ and $\{\hat{\rho}_{kn}(\alpha), \hat{\tilde{P}}_{1kn}(\alpha), \hat{\tilde{f}}_{kn}(\alpha)\}$ denote the same for $\hat{\mathcal{S}}_{1n}(\alpha)$.

Theorem 6.1. *Suppose that* $\mathrm{E}\|X\|^4_{L^2(E)} < \infty$. *Then, as* $n \to \infty$

$$\sqrt{n}\left\{\hat{\tilde{P}}_{1kn}(\alpha) - \tilde{P}_{1k}(\alpha)\right\} \xrightarrow{d} \tilde{P}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{Q}_{1k}(\alpha) + \tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{P}_{1k}(\alpha) \tag{53}$$

where $\mathcal{G}_1(\alpha)$ *is as in (49) and*

$$\tilde{Q}_{1k}(\alpha) = \sum_{j \neq k} \frac{1}{\rho_j(\alpha) - \rho_k(\alpha)} \tilde{P}_{1j}(\alpha).$$

In the case that $\text{rank}(\tilde{P}_{1k}(\alpha)) = 1$,

$$\sqrt{n}\left\{\hat{\tilde{f}}_{kn}(\alpha) - \tilde{f}_k(\alpha)\right\} \xrightarrow{d} \tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{f}_k(\alpha). \tag{54}$$

Proof: For each $k \in \mathbb{Z}$, let Γ_k denote a circle that encloses the eigenvalue $\rho_k(\alpha)$ but no other eigenvalue eigenvalue of $\mathcal{S}_1(\alpha)$. It follows from developments in the appendix that

$$\sqrt{n}\left\{\hat{\tilde{P}}_{1kn}(\alpha) - \tilde{P}_{1k}(\alpha)\right\} = \frac{\sqrt{n}}{2\pi i}\oint_{\Gamma_k} R(z)(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha))R(z)H_n(z,\alpha)dz \tag{55}$$

with $H_n(z,\alpha) \equiv \sum_{j=0}^{\infty}\left\{(\mathcal{S}_1(\alpha) - \hat{\mathcal{S}}_{1n}(\alpha))R(z)\right\}^j$ and $R(z)$ the resolvent of $\mathcal{S}_1(\alpha)$. Now since the integrand in (55) can be expanded into

$$R(z)(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha))R(z)H_n(z,\alpha) = R(z)(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha))R(z) + M(z,\alpha)$$

with

$$\begin{aligned}
M(z,\alpha) &\equiv R(z)(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha))R(z)\sum_{j=1}^{\infty}\left\{(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha))R(z)\right\}^j \\
&= R(z)(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha))R(z)(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha))R(z) \\
&+ R(z)(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha))R(z)(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha))R(z)(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha))R(z) + \cdots \\
&= \mathcal{O}_P(n^{-1}).
\end{aligned}$$

It follows that

$$\sqrt{n}\left\{\hat{\tilde{P}}_{1kn}(\alpha) - \tilde{P}_{1k}(\alpha)\right\} = \frac{\sqrt{n}}{2\pi i}\oint_{\Gamma_k} R(z)(\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha))R(z)dz + \mathcal{O}_P(n^{-1/2}). \tag{56}$$

We may now focus attention on the lead term in (56). From Corollary 6.1 and the continuous mapping theorem it follows that

$$\sqrt{n}\left\{\hat{\tilde{P}}_{1kn}(\alpha) - \tilde{P}_{1k}(\alpha)\right\} \xrightarrow{d} \frac{1}{2\pi i}\oint_{\Gamma_k} R(z)\mathcal{G}_1(\alpha)R(z)dz. \tag{57}$$

To simplify the last expression we write

$$R(z) = \sum_{k=1}^{\infty}\frac{1}{\rho_k(\alpha) - z}\tilde{P}_{1k}(\alpha) + \mathcal{O}((\rho_k(\alpha) - z)^{-2})$$

and all but the lead term will vanish when the contour integral is taken due to (96). The integrand in (57) can then be simplified as

$$R(z)\mathcal{G}_1(\alpha)R(z) = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{(\rho_k(\alpha) - z)(\rho_j(\alpha) - z)}\tilde{P}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{P}_{1j}(\alpha). \tag{58}$$

Applying the Cauchy integral formula to (58) ensures that

$$\frac{1}{2\pi i}\oint_{\Gamma_k}\frac{dz}{(\rho_k(\alpha)-z)(\rho_j(\alpha)-z)}=\frac{1}{2\pi i}\oint_{\Gamma_k}\frac{(\rho_i(\alpha)-z)dz}{(\rho_k(\alpha)-z)(\rho_j(\alpha)-z)(\rho_i(\alpha)-z)}$$

and the only case where the integral is non-zero is when exactly one of $\rho_k(\alpha)$ or $\rho_j(\alpha)$ is not equal to $\rho_i(\alpha)$. When, for example, $\rho_k(\alpha)=\rho_i(\alpha)$ and $\rho_i(\alpha)\neq\rho_j(\alpha)$ we have

$$\frac{1}{2\pi i}\oint_{\Gamma_k}\frac{(\rho_i(\alpha)-z)}{(\rho_k(\alpha)-z)(\rho_j(\alpha)-z)(\rho_i(\alpha)-z)}dz=\frac{1}{(\rho_j(\alpha)-\rho_k(\alpha))}$$

and hence

$$\begin{aligned}\sqrt{n}\left\{\hat{\tilde{P}}_{1kn}(\alpha)-\tilde{P}_{1k}(\alpha)\right\} &\xrightarrow{d} \sum_{i=1}^{\infty}\sum_{j\neq k}\frac{\delta_{ik}}{(\rho_j(\alpha)-\rho_k(\alpha))}\tilde{P}_{1i}(\alpha)\mathcal{G}_1(\alpha)\tilde{P}_{1j}(\alpha)\\ &+ \sum_{j=1}^{\infty}\sum_{i\neq k}\frac{\delta_{jk}}{(\rho_i(\alpha)-\rho_k(\alpha))}\tilde{P}_{1i}(\alpha)\mathcal{G}_1(\alpha)\tilde{P}_{1j}(\alpha)\\ &= \tilde{P}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{Q}_{1k}(\alpha)+\tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{P}_{1k}(\alpha)\end{aligned}$$

which establishes (53).

To obtain the limiting distribution of $\sqrt{n}\left\{\hat{\tilde{f}}_{kn}(\alpha)-\tilde{f}_k(\alpha)\right\}$, first observe that an application of Theorem 10.1 ensures that for large n and probability tending to 1, $\mathrm{rank}(\hat{\tilde{P}}_{1kn}(\alpha))=1$. Thus, we may write $\hat{\tilde{P}}_{1kn}(\alpha)=\left[\hat{\tilde{f}}_{kn}(\alpha)\otimes_{\mathcal{H}(K_1)}\hat{\tilde{f}}_{kn}(\alpha)\right]$ and hence

$$\begin{aligned}&\langle\hat{\tilde{P}}_{1kn}(\alpha)-\tilde{P}_{1k}(\alpha),\tilde{P}_{1k}(\alpha)\rangle_{HS}=\langle\hat{\tilde{P}}_{1kn}(\alpha),\tilde{P}_{1k}(\alpha)\rangle_{HS}-1\\ =\ &\langle(\hat{\tilde{f}}_{kn}(\alpha)\otimes_{\mathcal{H}(K_1)}\hat{\tilde{f}}_{kn}(\alpha)),(\tilde{f}_k(\alpha)\otimes_{\mathcal{H}(K_1)}\tilde{f}_k(\alpha))\rangle_{HS}-1\\ =\ &\langle\hat{\tilde{f}}_{kn}(\alpha),\tilde{f}_k(\alpha)\rangle^2_{\mathcal{H}(K_1)}-1\\ =\ &\left(\langle\hat{\tilde{f}}_{kn}(\alpha)-\tilde{f}_k(\alpha),\tilde{f}_k(\alpha)\rangle_{\mathcal{H}(K_1)}\right)\left(\langle\hat{\tilde{f}}_{kn}(\alpha),\tilde{f}_k(\alpha)\rangle_{\mathcal{H}(K_1)}+1\right).\end{aligned}$$

Furthermore, we note that

$$\begin{aligned}\sqrt{n}\left\{\hat{\tilde{f}}_{kn}(\alpha)-\tilde{f}_k(\alpha)\right\} &= \sqrt{n}\left[\tilde{P}_{1k}(\alpha)\right]\left\{\hat{\tilde{f}}_{kn}(\alpha)-\tilde{f}_k(\alpha)\right\}\\ &+ \sqrt{n}\left[I-\tilde{P}_{1k}(\alpha)\right]\left\{\hat{\tilde{f}}_{kn}(\alpha)-\tilde{f}_k(\alpha)\right\}.\end{aligned}\tag{59}$$

Focussing on the first term in the right hand side of (59) we see that

$$\begin{aligned}\sqrt{n}\quad&\left[\tilde{P}_{1k}(\alpha)\right]\left\{\hat{\tilde{f}}_{kn}(\alpha)-\tilde{f}_k(\alpha)\right\}\\ =\quad&\left[\sqrt{n}\langle\hat{\tilde{f}}_{kn}(\alpha)-\tilde{f}_k(\alpha),\tilde{f}_k(\alpha)\rangle_{\mathcal{H}(K_1)}\right]\tilde{f}_k(\alpha)\\ =\quad&\frac{\sqrt{n}\left(\langle\hat{\tilde{P}}_{1kn}(\alpha)-\tilde{P}_{1k}(\alpha),\tilde{P}_{1k}(\alpha)\rangle_{HS}\right)}{\left((\hat{\tilde{f}}_{kn}(\alpha),\tilde{f}_k(\alpha))_{\mathcal{H}(K_1)}+1\right)}\tilde{f}_k(\alpha).\end{aligned}\tag{60}$$

Due to the continuity of the inner product and (53) it follows that

$$\begin{aligned}
\sqrt{n} \quad & \langle \hat{\tilde{P}}_{1kn}(\alpha) - \tilde{P}_{1k}(\alpha), \tilde{P}_{1k}(\alpha) \rangle_{HS} \\
\xrightarrow{d} \quad & \langle \tilde{P}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{Q}_{1k}(\alpha), \tilde{P}_{1k}(\alpha) \rangle_{HS} + \langle \tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{P}_{1k}(\alpha), \tilde{P}_{1k}(\alpha) \rangle_{HS} \\
= \quad & \operatorname{tr}\left(\tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{P}_{1k}(\alpha)\right) + \operatorname{tr}\left(\tilde{P}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{Q}_{1k}(\alpha)\tilde{P}_{1k}(\alpha)\right) \\
= \quad & \operatorname{tr}\left(\tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{P}_{1k}(\alpha)\right) + 0 \\
= \quad & \operatorname{tr}\left(\tilde{P}_{1k}(\alpha)\tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\right) \\
= \quad & 0
\end{aligned} \tag{61}$$

because

$$\tilde{Q}_{1k}(\alpha)\tilde{P}_{1k}(\alpha) = \sum_{j \neq k} \frac{1}{\rho_j(\alpha) - \rho_k(\alpha)} \tilde{P}_{1j}(\alpha)\tilde{P}_{1k}(\alpha) = 0 = \tilde{P}_{1k}(\alpha)\tilde{Q}_{1k}(\alpha). \tag{62}$$

Consequently, the numerator in (60) converges in probability to 0 whereas the denominator $\left(\langle \hat{\tilde{f}}_{kn}(\alpha), \tilde{f}_k(\alpha) \rangle_{\mathcal{H}(K_1)} + 1\right) = (2 + \mathcal{O}_P(n^{-1/2}))$. Slutsky's theorem then implies that $\sqrt{n}\left[\tilde{P}_{1k}(\alpha)\right]\left\{\hat{\tilde{f}}_{kn}(\alpha) - \tilde{f}_k(\alpha)\right\} \xrightarrow{d} 0$ and hence $\langle \hat{\tilde{f}}_{kn}(\alpha), \tilde{f}_k(\alpha) \rangle_{\mathcal{H}(K_1)} \xrightarrow{p} 1$.

To address the second term on the right hand side of (59) we observe that as a consequence of Slutsky's theorem

$$\begin{aligned}
\sqrt{n} \quad & \left[I - \tilde{P}_{1k}(\alpha)\right]\left\{\hat{\tilde{f}}_{kn}(\alpha) - \tilde{f}_k(\alpha)\right\} \\
= \quad & \sqrt{n}\left[I - \tilde{P}_{1k}(\alpha)\right]\hat{\tilde{f}}_{kn}(\alpha) \\
= \quad & \frac{\sqrt{n}\left[I - \tilde{P}_{1k}(\alpha)\right]\left[\hat{\tilde{P}}_{1kn}(\alpha)\right]\tilde{f}_k(\alpha)}{\langle \hat{\tilde{f}}_{kn}(\alpha), \tilde{f}_k(\alpha) \rangle_{\mathcal{H}(K_1)}} \\
= \quad & \frac{\sqrt{n}\left[I - \tilde{P}_{1k}(\alpha)\right]\left[\hat{\tilde{P}}_{1kn}(\alpha) - \tilde{P}_{1k}(\alpha)\right]\tilde{f}_k(\alpha)}{\langle \hat{\tilde{f}}_{kn}(\alpha), \tilde{f}_k(\alpha) \rangle_{\mathcal{H}(K_1)}} \\
\xrightarrow{d} \quad & \left[I - \tilde{P}_{1k}(\alpha)\right]\left[\tilde{P}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{Q}_{1k}(\alpha) + \tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{P}_{1k}(\alpha)\right]\tilde{f}_k(\alpha) \\
= \quad & \left[I - \tilde{P}_{1k}(\alpha)\right]\left[\tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\right]\tilde{f}_k(\alpha) \\
= \quad & \tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{f}_k(\alpha).
\end{aligned} \tag{63}$$

Equation (63) establishes (54) which completes the proof. $\diamondsuit$

We may now derive the limiting distribution for $\sqrt{n}\,[\hat{\rho}_{kn}(\alpha) - \rho_k(\alpha)]$, where $\{\hat{\rho}_{kn}(\alpha), \rho_k(\alpha)\}$ denotes the k^{th} distinct eigenvalue associated with $\{\hat{\mathcal{S}}_{1n}(\alpha), \mathcal{S}_1(\alpha)\}$. In the following result, if $\rho_k(\alpha)$ has geometric multiplicity d_k then $\sqrt{n}\,[\hat{\rho}_{kn}(\alpha) - \rho_k(\alpha)]$ will be regarded as a vector of dimension d_k.

Theorem 6.2. *Assume that* $\mathrm{E}\|X\|^4_{L^2(E)} < \infty$ *and the* k^{th} *regularized canonical correlation,* $\rho_k(\alpha)$*, has geometric multiplicity* d_k*. Then,*

$$\begin{aligned}
\sqrt{n}\left[\hat{\rho}_{kn}(\alpha) - \rho_k(\alpha)\right] &= \sqrt{n}\left[\hat{\tilde{P}}_{1kn}(\alpha)\hat{\mathcal{S}}_{1n}(\alpha)\hat{\tilde{P}}_{1kn}(\alpha) - \rho_k(\alpha)\tilde{P}_{1k}(\alpha)\right] \\
&\xrightarrow{d} \tilde{P}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{P}_{1k}(\alpha)
\end{aligned}$$

with $\mathcal{G}_1(\alpha)$ *the Gaussian random variable in (50). Furthermore,* $\tilde{P}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{P}_{1k}(\alpha)$ *has dimension* d_k *and, in the special case that* $d_k = 1$*,*

$$\sqrt{n}\left(\hat{\rho}_{kn}(\alpha) - \rho_k(\alpha)\right) \xrightarrow{d} N(0, \sigma_{kk}(\alpha))$$

where $N(0, \sigma_{kk}(\alpha))$ *denotes a normal distribution with zero mean and variance*

$$\sigma_{kk}(\alpha) = \mathrm{E}\left[\langle \tilde{f}_k(\alpha), \mathcal{G}_1(\alpha)\tilde{f}_k(\alpha)\rangle^2_{\mathcal{H}(K_1)}\right].$$

Proof: Let $\rho_k(\alpha)$ denote the k^{th} distinct eigenvalue of $\mathcal{S}_1(\alpha)$ and assume that it has multiplicity d_k. As $\|\hat{\tilde{P}}_{1kn}(\alpha) - \tilde{P}_{1k}(\alpha)\| \xrightarrow{p} 0$, Theorem 10.1 ensures that for n large enough, $\mathrm{rank}(\hat{\tilde{P}}_{1kn}(\alpha)) = \mathrm{rank}(\tilde{P}_{1k}(\alpha)) = d_k$ with probability tending to 1. Now observe that

$$\begin{aligned}
\sqrt{n}\left[\hat{\rho}_{kn}(\alpha) - \rho_k(\alpha)\right] &= \sqrt{n}\left[\hat{\tilde{P}}_{1kn}(\alpha)\hat{\mathcal{S}}_{1n}(\alpha)\hat{\tilde{P}}_{1kn}(\alpha) - \rho_k(\alpha)\tilde{P}_{1k}(\alpha)\right] \\
&= \sqrt{n}\left[\sum_{j=1}^{3}\hat{\mathcal{B}}_{kj}(\alpha)\right]
\end{aligned}$$

where

$$\begin{aligned}
\hat{\mathcal{B}}_{k1}(\alpha) &\equiv [\hat{\tilde{P}}_{1kn}(\alpha) - \tilde{P}_{1k}(\alpha)]\hat{\mathcal{S}}_{1n}(\alpha)\hat{\tilde{P}}_{1kn}(\alpha), \\
\hat{\mathcal{B}}_{k2}(\alpha) &\equiv \tilde{P}_{1k}(\alpha)[\hat{\mathcal{S}}_{1n}(\alpha) - \mathcal{S}_1(\alpha)]\hat{\tilde{P}}_{1kn}(\alpha), \\
\hat{\mathcal{B}}_{k3}(\alpha) &\equiv \tilde{P}_{1k}(\alpha)\mathcal{S}_1(\alpha)[\hat{\tilde{P}}_{1kn}(\alpha) - \tilde{P}_{1k}(\alpha)].
\end{aligned}$$

Equations (62), (53) and Slutsky's theorem then ensure that

$$\begin{aligned}
&\|\hat{\mathcal{B}}_{k1}(\alpha)\|^2_{HS} \xrightarrow{d} \|\left[\tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{P}_{1k}(\alpha) + \tilde{P}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{Q}_{1k}(\alpha)\right]\mathcal{S}_1(\alpha)\tilde{P}_{1k}(\alpha)\|^2_{HS} \\
\leq\ & \|\tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\mathcal{S}_1(\alpha)\tilde{P}_{1k}(\alpha)\|^2_{HS} + \|\tilde{P}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{Q}_{1k}(\alpha)\mathcal{S}_1(\alpha)\tilde{P}_{1k}(\alpha)\|^2_{HS} \\
=\ & \|\tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\mathcal{S}_1(\alpha)\tilde{P}_{1k}(\alpha)\|^2_{HS} + \|\tilde{P}_{1k}(\alpha)\mathcal{G}_1(\alpha)\left[\tilde{Q}_{1k}(\alpha)\tilde{P}_{1k}(\alpha)\right]\mathcal{S}_1(\alpha)\|^2_{HS} \\
=\ & \mathrm{tr}\left(\tilde{P}_{1k}(\alpha)\mathcal{S}_1(\alpha)\mathcal{G}_1(\alpha)\tilde{Q}^2_{1k}(\alpha)\mathcal{G}_1(\alpha)\mathcal{S}_1(\alpha)\tilde{P}_{1k}(\alpha)\right) + 0 \\
=\ & \mathrm{tr}\left(\mathcal{S}_1(\alpha)\left[\tilde{P}_{1k}(\alpha)\tilde{Q}_{1k}(\alpha)\right]\mathcal{G}_1(\alpha)\mathcal{G}_1(\alpha)\left[\tilde{Q}_{1k}(\alpha)\tilde{P}_{1k}(\alpha)\right]\mathcal{S}_1(\alpha)\right) = 0.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\|\hat{\mathcal{B}}_{k3}(\alpha)\|_{HS}^2 \xrightarrow{d} \|\tilde{P}_{1k}(\alpha)\mathcal{S}_1(\alpha)\left[\tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{P}_{1k}(\alpha)+\tilde{P}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{Q}_{1k}(\alpha)\right]\|_{HS}^2\\
\leq\ & \|\mathcal{S}_1(\alpha)\left[\tilde{P}_{1k}(\alpha)\tilde{Q}_{1k}(\alpha)\right]\mathcal{G}_1(\alpha)\tilde{P}_{1k}(\alpha)\|_{HS}^2+\|\tilde{P}_{1k}(\alpha)\mathcal{S}_1(\alpha)\mathcal{G}_1(\alpha)\tilde{Q}_{1k}(\alpha)\|_{HS}^2\\
=\ & 0+\mathrm{tr}\left(\tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\mathcal{S}_1(\alpha)\tilde{P}_{1k}(\alpha)\mathcal{S}_1(\alpha)\mathcal{G}_1(\alpha)\tilde{Q}_{1k}(\alpha)\right)\\
=\ & \mathrm{tr}\left(\mathcal{G}_1(\alpha)\left[\tilde{Q}_{1k}(\alpha)\tilde{P}_{1k}(\alpha)\right]\mathcal{S}_1(\alpha)\tilde{P}_{1k}(\alpha)\mathcal{S}_1(\alpha)\left[\tilde{P}_{1k}(\alpha)\tilde{Q}_{1k}(\alpha)\right]\mathcal{G}_1(\alpha)\right)=0.
\end{aligned}$$

Hence Corollary 6.1 and Slutsky's Theorem ensure that

$$\sqrt{n}\hat{\mathcal{B}}_{k2}(\alpha)\xrightarrow{d}\tilde{P}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{P}_{1k}(\alpha)$$

which proves the first part of the theorem.

To see the validity of the second part of the theorem, assume that $d_k=1$ and observe that

$$\sqrt{n}\left(\hat{\rho}_{kn}(\alpha)-\rho_k(\alpha)\right)=\sqrt{n}\left\{\sum_{j=1}^{3}\hat{\mathcal{C}}_{kj}(\alpha)\right\}$$

where

$$\begin{aligned}
\hat{\mathcal{C}}_{k1}(\alpha)&\equiv\langle[\hat{\tilde{f}}_{kn}(\alpha)-\tilde{f}_k(\alpha)],\hat{\mathcal{S}}_{1n}(\alpha)\hat{\tilde{f}}_{kn}(\alpha)\rangle_{\mathcal{H}(K_1)},\\
\hat{\mathcal{C}}_{k2}(\alpha)&\equiv\langle\tilde{f}_k(\alpha),[\hat{\mathcal{S}}_{1n}(\alpha)-\mathcal{S}_1(\alpha)]\hat{\tilde{f}}_{kn}(\alpha)\rangle_{\mathcal{H}(K_1)},\\
\hat{\mathcal{C}}_{k3}(\alpha)&\equiv\langle\tilde{f}_k(\alpha),\mathcal{S}_1(\alpha)[\hat{\tilde{f}}_{kn}(\alpha)-\tilde{f}_k(\alpha)]\rangle_{\mathcal{H}(K_1)}.
\end{aligned}$$

Note that $\hat{\mathcal{C}}_{k1}(\alpha)\xrightarrow{p}0$ and $\hat{\mathcal{C}}_{k3}(\alpha)\xrightarrow{p}0$ as a consequence of equations (54), (62), and Slutsky's theorem since

$$\begin{aligned}
\hat{\mathcal{C}}_{k1}(\alpha)&\xrightarrow{d}\langle\tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{f}_k(\alpha),\mathcal{S}_1(\alpha)\tilde{f}_k(\alpha)\rangle_{\mathcal{H}(K_1)}\\
&=\sum_{j\neq k}\frac{\rho_k(\alpha)}{(\rho_j(\alpha)-\rho_k(\alpha))}\langle\tilde{P}_{1j}(\alpha)\mathcal{G}_1(\alpha)\tilde{f}_k(\alpha),\tilde{f}_k(\alpha)\rangle_{\mathcal{H}(K_1)}=0
\end{aligned}$$

and,

$$\begin{aligned}
\hat{\mathcal{C}}_{k3}(\alpha)&\xrightarrow{d}\langle\tilde{f}_k(\alpha),\mathcal{S}_1(\alpha)\tilde{Q}_{1k}(\alpha)\mathcal{G}_1(\alpha)\tilde{f}_k(\alpha)\rangle_{\mathcal{H}(K_1)}\\
&=\sum_{j\neq k}\frac{\rho_k(\alpha)}{(\rho_j(\alpha)-\rho_k(\alpha))}\langle\tilde{P}_{1j}(\alpha)\tilde{f}_k(\alpha),\mathcal{G}_1(\alpha)\tilde{f}_k(\alpha)\rangle_{\mathcal{H}(K_1)}=0.
\end{aligned}$$

Application of Theorem 6.1 and Slutsky's Theorem then ensures that $\hat{\mathcal{C}}_{k2}(\alpha)\xrightarrow{d}\langle\tilde{f}_k(\alpha),\mathcal{G}_1(\alpha)\tilde{f}_k(\alpha)\rangle_{\mathcal{H}(K_1)}$. Which completes the proof since

$$\mathrm{E}\left[\langle\tilde{f}_k(\alpha),\mathcal{G}_1(\alpha)\tilde{f}_k(\alpha)\rangle_{\mathcal{H}(K_1)}\right]=0$$

and hence

$$\text{Var}\left[\langle \tilde{f}_k(\alpha), \mathcal{G}_1(\alpha)\tilde{f}_k(\alpha)\rangle_{\mathcal{H}(K_1)}\right] = \text{E}\left[\langle \tilde{f}_k(\alpha), \mathcal{G}_1(\alpha)\tilde{f}_k(\alpha)\rangle^2_{\mathcal{H}(K_1)}\right] \equiv \sigma_{kk}(\alpha).$$

$\diamondsuit$

Now as the Gaussian operator $\mathcal{G}_1(\alpha)$ is Hilbert-Schmidt, we note that the variances $\sigma_{kk}(\alpha) \downarrow 0$ as $k \uparrow \infty$.

Of natural interest is the degree of correlation between the regularized correlation estimators $\{\hat{\rho}_{kn}(\alpha), \hat{\rho}_{jn}(\alpha)\}$ with $j \neq k$. To investigate this association let us take the simple case where for $j \neq k$, the multiplicities $d_k = d_j = 1$. Then,

$$\begin{aligned}
& \sigma_{jk}(\alpha) \equiv \text{Cov}[\hat{\rho}_{kn}(\alpha), \hat{\rho}_{jn}(\alpha)] \\
= \; & \text{E}\left[\langle \hat{\tilde{f}}_{kn}(\alpha), \hat{\mathcal{S}}_{1n}(\alpha)\hat{\tilde{f}}_{kn}(\alpha)\rangle_{\mathcal{H}(K_1)}\langle \hat{\tilde{f}}_{jn}(\alpha), \hat{\mathcal{S}}_{1n}(\alpha)\hat{\tilde{f}}_{jn}(\alpha)\rangle_{\mathcal{H}(K_1)}\right] \\
= \; & \text{E}\left[\langle(\hat{\tilde{f}}_{kn}(\alpha) \otimes_{\mathcal{H}(K_1)} \hat{\tilde{f}}_{kn}(\alpha)), \left[\hat{\mathcal{S}}_{1n}(\alpha) \otimes_{HS} \hat{\mathcal{S}}_{1n}(\alpha)\right](\hat{\tilde{f}}_{jn}(\alpha) \otimes_{\mathcal{H}(K_1)} \hat{\tilde{f}}_{jn}(\alpha))\rangle_{HS}\right] \\
= \; & \langle(\tilde{f}_k(\alpha) \otimes_{\mathcal{H}(K_1)} \tilde{f}_k(\alpha)), \text{E}\left[\hat{\mathcal{S}}_{1n}(\alpha) \otimes_{HS} \hat{\mathcal{S}}_{1n}(\alpha)\right](\tilde{f}_j(\alpha) \otimes_{\mathcal{H}(K_1)} \tilde{f}_j(\alpha))\rangle_{HS} \\
= \; & \langle(\tilde{f}_k(\alpha) \otimes_{\mathcal{H}(K_1)} \tilde{f}_k(\alpha)), \Sigma_1(\alpha)(\tilde{f}_j(\alpha) \otimes_{\mathcal{H}(K_1)} \tilde{f}_j(\alpha))\rangle_{HS_1} \equiv [\Sigma_1(\alpha)]_{jk}
\end{aligned}$$

where $\Sigma_1(\alpha) \equiv \text{E}\left[\hat{\mathcal{S}}_{1n}(\alpha) \otimes_{HS} \hat{\mathcal{S}}_{1n}(\alpha)\right]$, and $[\Sigma_1(\alpha)]_{jk}$ is the $\{j,k\}^{th}$ element of $\Sigma_1(\alpha)$. These developments suggest that the j^{th} and k^{th} regularized canonical correlation estimators are not necessarily independent.

There are many similarities between the Tikhinov regularized version of canonical correlation analysis discussed here and those discussed in Cupidon et al. [6]. However, it is important to distinctions between Cupidon et al. [6] and the method discussed here. Firstly, in Cupidon et al [6] the regularized operators discussed were of the form $(S_1+\alpha I)^{-1/2}S_{12}(S_2+\alpha I)^{-1}S_{21}(S_1+\alpha I)^{-1/2}$ whereas in our approach they are $S_{12}(S_2+\alpha I)^{-1}S_{21}(S_1+\alpha I)^{-1}$. Secondly, in Cupidon et al. [6] the operator approaches RR^* as the regularization parameter approaches zero, and in our approach it tends to TT^*, an RKHS based operator which has well posed solutions on a closed domain. Finally, the variance in the asymptotic distribution of $\hat{S}_{12}(\hat{S}_2+\alpha I)^{-1}\hat{S}_{21}(\hat{S}_1+\alpha I)^{-1}$ is the sum of 4 terms whereas the variance of $(\hat{S}_1+\alpha I)^{-1/2}\hat{S}_{12}(\hat{S}_2+\alpha I)^{-1}\hat{S}_{21}(\hat{S}_1+\alpha I)^{-1/2}$ involves 5 terms.

7. TSVD Regularization Approach

In the Tikhinov approach to regularization, the operators $\{S_1, S_2\}$ are replaced with the operators $\{(S_1+\alpha I), (S_2+\alpha I)\}$ to obtain invertible operators. By contrast, the truncated singular value decomposition (TSVD) method of regularization replaces the compact operators $\{S_1, S_2\}$ with the finite rank (and rank-deficient) operators

$$\begin{aligned}
S_1(\alpha) &= \sum_{\lambda_{1i}>\alpha} \lambda_{1i}\phi_{1i} \otimes_{L^2(E_1)} \phi_{1i} \text{ and} \\
S_2(\alpha) &= \sum_{\lambda_{2i}>\alpha} \lambda_{2i}\phi_{2i} \otimes_{L^2(E_1)} \phi_{2i}.
\end{aligned}$$

Let us now define $m_1(\alpha) = \{\# \text{ of } \lambda_{1i} > \alpha\}$ with similar definition holding for $m_2(\alpha)$. To ensure the equal dimensionality of the truncated versions of S_1 and S_2, it is advantageous to re-parameterize TSVD regularization in terms of $\{m_1(\alpha), m_2(\alpha)\}$, rather than α. In this regard, for simplicity we will always take $m = m_1(\alpha) = m_2(\alpha)$. Notice that, under this re-parametrization, the compact operators S_1 and S_2 are replaced by the finite dimensional operators $S_1(m) \equiv S_1\Pi_1(m)$ and $S_2(m) \equiv S_2\Pi_2(m)$, where $\Pi_1(m) \equiv \sum_{i=1}^m P_{1i}$ and $\Pi_2(m) \equiv \sum_{i=1}^m P_{2i}$, are the projection operators associated with the largest m eigenvalues of S_1 and S_2 (or cumulative projection operators). Much like α in Tikhinov regularization, the truncation parameter m is the regularization parameter. In TSVD we are interested in the case where $m = m(\alpha) \to \infty$ which occurs when $\alpha \downarrow 0$. However, it is important to mention that TSVD regularization is widely used in statistical practice, as it is common for a practicing statistician to discard right and left eigenvectors corresponding to small singular values after looking at, for example, a scree plot of the singular values.

Development of the theory associated with the TSVD version of regularized canonical correlation analysis can now proceed along lines that are parallel to the developments in Sections 5 and 6. Accordingly, let us define the operators $R(m) : \mathcal{H}_2 \mapsto \mathcal{H}_1$ and $T(m) : \mathcal{H}(K_2) \mapsto \mathcal{H}(K_1)$ by

$$R(m) \equiv S_1(m)^{1/2\dagger} S_{12} S_2(m)^{1/2\dagger} \tag{64}$$

and

$$T(m) \equiv \Gamma_1 S_1(m)^{1/2\dagger} S_{12} S_2(m)^{1/2\dagger} \Gamma_2^{-1}. \tag{65}$$

As all operators in (64) and (65) are bounded and S_{12} is Hilbert-Schmidt, both $R(m)$ and $T(m)$ are Hilbert-Schmidt. Also, as the regularization parameter $m \to \infty$, both $\Pi_1(m)$ and $\Pi_2(m)$ converge to the identity. Thus, by Theorem 4.1, $R(m)$ converges in operator norm to the operator $\Gamma_1^{-1} T \Gamma_2$ and the continuity of Γ_1 and Γ_2 then entail that

$$T(m) = \Gamma_1 R(m) \Gamma_2^{-1} \to T \quad \text{as} \quad m \to \infty$$

with convergence in operator norm.

Now suppose that $\{\rho_j(m), f_j(m), g_j(m)\}_{j=1}^{\operatorname{rank}(R(m)^* R(m))}$ is the singular system for $R(m)$ with

$$R(m) = \sum_{i,j=1}^{m} \frac{\gamma_{ij}}{\sqrt{\lambda_{1i}\lambda_{2j}}} \phi_{2j} \otimes_{\mathcal{H}_2} \phi_{1i} = \sum_{j=1}^{\operatorname{rank}(R(m)^* R(m))} \rho_j(m) \left[g_j(m) \otimes_{\mathcal{H}_2} f_j(m) \right].$$

Then,

$$\begin{aligned} T(m) = \Gamma_1 R(m) \Gamma_2^{-1} &= \sum_{j=1}^{\operatorname{rank}(R(m)^* R(m))} \rho_j(m) \left[(\Gamma_2 g_j(m)) \otimes_{\mathcal{H}(K_2)} (\Gamma_1 f_j(m)) \right] \\ &= \sum_{j=1}^{\operatorname{rank}(T(m)^* T(m))} \rho_j(m) \left[\tilde{g}_j(m) \otimes_{\mathcal{H}(K_2)} \tilde{f}_j(m) \right] \end{aligned}$$

with $\{\tilde{f}_j(m) = \Gamma_1 f_j(m), \tilde{g}_j(m) = \Gamma_2 g_j(m)\}_{j=1}^{\text{rank}(T(m)^*T(m))}$. We note that in general $\text{rank}(T(m)^*T(m)) \le m$. Now by (16) and because $\{\phi_{ij}\}_{j=1}^{m}$ are CONSs for $[\Pi_i(m)\mathcal{H}_i]$, for $i = 1, 2$, it follows that the canonical weight functions in $\mathcal{H}(K_1)$ and $\mathcal{H}(K_2)$ may be written by

$$\tilde{f}_j(m) = \sum_{i=1}^{m} \lambda_{1i} f_{ji}(m)\phi_{1i} \quad \text{and} \quad \tilde{g}_j(m) = \sum_{i=1}^{m} \lambda_{2i} g_{ji}(m)\phi_{2i}$$

with

$$f_{ji}(m) = \langle f_j(m), \phi_{1i}\rangle_{\mathcal{H}_1} \quad \text{and} \quad g_{ji}(m) = \langle g_j(m), \phi_{2j}\rangle_{\mathcal{H}_2}.$$

The corresponding regularized canonical variables in $L^2_{X_1}$ and $L^2_{X_2}$ are

$$\begin{aligned} U_j(m) &= \Psi_1(\tilde{f}_j(m)) = \textstyle\sum_{i=1}^{m} f_{ji}(m)\langle X_1, \phi_{1i}\rangle_{\mathcal{H}_2} \text{ and} \\ V_j(m) &= \Psi_2(\tilde{g}_j(m)) = \textstyle\sum_{i=1}^{m} g_{ji}(m)\langle X_2, \phi_{2i}\rangle_{\mathcal{H}_2}. \end{aligned}$$

The TSVD parallel to Theorems 5.1 and 5.2 from Cupidon et al. [6] also hold.

Theorem 7.1. *For any $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$, with $\tilde{f} = \Gamma_1(f)$, $\tilde{g} = \Gamma_2(g)$*

$$0 \le \rho_j^2(m) \uparrow \rho_j^2 \le 1, \text{ as } m \to \infty. \tag{66}$$

Proof: The convergence is from below since

$$\|T(m)\| = \|\Pi_1(m)T\Pi_2(m)\| < \|\Pi_1(m)\|\ \|T\|\ \|\Pi_2(m)\| < \|T\|.$$

To see that (66) holds, fix $j \ge 1$ and observe that as $m \uparrow \infty$

$$\left(\rho_j^2 - \rho_j^2(m)\right) = |\rho_j^2 - \rho_j^2(m)| \le \|TT^* - T(m)T^*(m)\| \downarrow 0.$$

$\diamondsuit$

Theorem 7.2. *Let $\{\tilde{f}_j(m), \tilde{g}_j(m)\} = \{\Gamma_1(f_j(m)), \Gamma_2(g_j(m))\} \in \{\mathcal{H}(K_1), \mathcal{H}(K_2)\}$ denote the regularized weight functions corresponding to the TSVD version of canonical correlation analysis. Then, as $m \to \infty$ for $j = 1, 2, \ldots,$*

$$\|\tilde{f}_j(m) - \tilde{f}_j\|_{\mathcal{H}(K_1)} \to 0 \quad \textit{and } \|\tilde{g}_j(m) - \tilde{g}_j\|_{\mathcal{H}(K_2)} \to 0.$$

Proof: The proof here parallels the one for Theorem 5.2. The idea is that since $\|T(m)T^*(m) - TT^*\| \to 0$ as $m \to \infty$, this implies that for any $j \in \mathbb{Z}$, the corresponding eigenprojection operators $\|P_j(m) - P_j\|_{HS} \to 0$. If we now assume, WLOG, that $\langle \tilde{f}_j(m), \tilde{f}_j\rangle_{\mathcal{H}(K_1)} \ge 0$, the relation

$$\begin{aligned} \|P_j(m) - P_j\|^2_{HS} &= \langle P_j(m) - P_j, P_j(m) - P_j\rangle_{HS} \\ &= 2 - 2\langle P_j(m), P_j\rangle_{HS} \\ &= 2 - 2\langle (\tilde{f}_j(m) \otimes_{\mathcal{H}(K_1)} \tilde{f}_j(m)), (\tilde{f}_j \otimes_{\mathcal{H}(K_1)} \tilde{f}_j)\rangle_{HS} \\ &= 2 - 2\langle \tilde{f}_j(m), \tilde{f}_j\rangle^2_{\mathcal{H}(K_1)} \\ &= 2(1 - \langle \tilde{f}_j(m), \tilde{f}_j\rangle_{\mathcal{H}(K_1)})(1 + \langle \tilde{f}_j(m), \tilde{f}_j\rangle_{\mathcal{H}(K_1)}) \\ &= \|\tilde{f}_j(m) - \tilde{f}_j\|^2_{\mathcal{H}(K_1)}(1 + \langle \tilde{f}_j(m), \tilde{f}_j\rangle_{\mathcal{H}(K_1)}) \\ &\ge \|\tilde{f}_j(m) - \tilde{f}_j\|^2_{\mathcal{H}(K_1)} \end{aligned}$$

implies that

$$\|\tilde{f}_j(m) - \tilde{f}_j\|^2_{\mathcal{H}(K_1)} \leq \|P_j(m) - P_j\|_{HS} \to 0$$

as the regularization parameter $m \uparrow \infty$. $\diamondsuit$

Theorem 7.2, along with the continuity of the mappings Ψ_1 and Ψ_2, ensures the convergence of the regularized canonical variables $U_j(m) = \Psi_1(\tilde{f}_j(m))$ and $V_j(m) = \Psi_2(\tilde{g}_j(m))$ to the true canonical variables $\Psi_1(\tilde{f}_j)$ and $\Psi_2(\tilde{g}_j)$, as $m \to \infty$.

Let us now discuss the computation of the singular value decomposition of $T(m)$. To accomplish this it suffices to consider the eigenvalue-eigenvector decomposition of $T(m)T^*(m)$. This is the finite rank operator given by

$$\begin{aligned} \mathcal{T}_1(m) &\equiv T(m)T^*(m) = \Gamma_1 R(m) R^*(m) \Gamma_1^{-1} \\ &= \Gamma_1 \Pi_1(m) S_1(m)^{1/2\dagger} S_{12} S_2(m)^{-1} S_{21} S_1(m)^{1/2\dagger} \Pi_1(m) \Gamma_1^{-1}. \quad (67) \end{aligned}$$

As was true for the Tikhinov case, problems arise from the presence of the unknown Γ_1 in $T(m)$. However, unlike Tikhinov regularization the operators involved are finite rank. Note that in the finite rank case $\mathrm{Im}(S_1^{1/2}) = \overline{\mathrm{Im}(S_1^{1/2})} = \ker(S_1)^\perp$ and we may substitute $\Gamma_1 \Pi_1(m)$ with $S_1^{1/2}\Pi_1(m)$ directly. Upon direct substitution of $S_1^{1/2}$ for Γ_1 in (67) we obtain the operator

$$\mathcal{S}_1(m) = \Pi_1(m) S_{12} S_2(m)^\dagger S_{21} S_1(m)^\dagger \quad (68)$$

which is a mapping from $\mathcal{H}_1$ into $\mathcal{H}_2$. Much like its Tikhinov cousin, the operator $\mathcal{S}_1(m)$ is self-adjoint since

$$\begin{aligned} \mathcal{S}_1(m) &= \sum_{i,j=1}^{m} \left(\frac{\gamma_{ij}^2}{\lambda_{1i}\lambda_{2j}} \right) [\phi_{1i} \otimes_{\mathcal{H}_1} \phi_{1j}] \\ &= S_1(m)^\dagger S_{12} S_2(m)^\dagger S_{21} \Pi_1(m) = \mathcal{S}_1^*(m). \end{aligned}$$

Additionally, since the operator is finite rank, it is Hilbert-Schmidt and admits the eigenvalue-eigenvector decomposition

$$\mathcal{S}_1(m) = \sum_{j=1}^{\mathrm{rank}(\mathcal{S}_1(m))} \rho_j^2(m) \, [f_j(m) \otimes_{\mathcal{H}_1} f_j(m)]$$

with $\{\rho_j^2(m), f_j(m)\}_{j=1}^{\mathrm{rank}(\mathcal{S}_1(m))}$, the eigensystem for $\mathcal{S}_1(m)$.

All of the themes discussed in Section 2.1 are still applicable here. For example, since the eigenfunctions $\{f_j(m)\}$ are in $\mathrm{Im}(S^{1/2})$, they belong to both $\mathcal{H}(K_1)$ and $\mathcal{H}_1$. If the eigenfunctions are considered to be elements of $\mathcal{H}(K_1)$ we will notate these as $\{\tilde{f}_j(m)\}$ with $\{\tilde{f}_j(m) = f_j(m)\}$. We will now show that as $m \uparrow \infty$

$$\Gamma_1^{-1} \mathcal{S}_1(m) \Gamma_1 \to \Gamma_1^{-1} T T^* \Gamma_1.$$

To see this notice that

$$
\begin{aligned}
\Gamma_1^{-1}\mathcal{S}_1(m)\Gamma_1 \quad &\equiv \quad \sum_{i,j=1}^{m} \frac{\gamma_{ij}^2}{\lambda_{1i}\lambda_{2j}} \left[(\tilde{\phi}_{1i}\Gamma_1) \otimes_{\mathcal{H}_1} (\Gamma_1^{-1}\tilde{\phi}_{1j})\right] \\
&\rightarrow \quad \sum_{i,j=1}^{\infty} \frac{\gamma_{ij}^2}{(\lambda_{1i})(\lambda_{2j})} \left[\phi_{1i} \otimes_{\mathcal{H}_1} \phi_{1j}\right] = \Gamma_1^{-1}TT^*\Gamma_1.
\end{aligned}
$$

Therefore, since $\mathcal{S}_1(m)$ is Hilbert-Schmidt, Theorem 4.1 ensures that as the regularization parameter $m \to \infty$,

$$
\|\Gamma_1^{-1}\mathcal{S}_1(m)\Gamma_1 - \Gamma_1^{-1}TT^*\Gamma_1\|_{\mathcal{H}_1} = \|\mathcal{S}_1(m) - TT^*\|_{\mathcal{H}(K_1)} \to 0.
$$

If $\mathcal{S}_1(m)$ is regarded as an operator on $\mathcal{H}(K_1)$, we let $\{\rho_j^2(m), \tilde{f}_j(m)\}$ denote the eigenvalue and eigenvector pairs for the operator and using this notation

$$
\mathcal{S}_1(m) = \sum_{j=1}^{\operatorname{rank}(\mathcal{S}_1(m))} \rho_j^2(m) \left[\tilde{f}_j(m) \otimes_{\mathcal{H}(K_1)} \tilde{f}_j(m)\right].
$$

8. Asymptotics for the TSVD Operators

In this section we will discuss the large sample distribution and consistency of sample versions of $\mathcal{S}_1(m)$. The obvious estimator for this quantity is given by

$$
\hat{\mathcal{S}}_{1n}(m) \equiv \hat{\Pi}_{1n}(m)\hat{S}_{12n}\hat{S}_{2n}(m)^{\dagger}\hat{S}_{21n}\hat{S}_{1n}(m)^{\dagger} \tag{69}
$$

with $\hat{\Pi}_{in}(m) \equiv \sum_{j=1}^{m} \hat{P}_{ijn}$ and $\hat{S}_{in}(m)^{\dagger} = (\hat{S}_{in}\hat{\Pi}_{in}(m))^{\dagger} = (\hat{S}_{in})^{\dagger}\hat{\Pi}_{in}(m)$ for $i = 1, 2$. By considering each factor associated with the TSVD operators in equation (69) it is clear that the asymptotic distribution will differ from its corresponding Tikhinov counterpart. We begin our analysis by assuming that the joint process X has zero mean and $\mathrm{E}[\|X\|_{L^2(E)}^4] < \infty$. Accordingly, we need to develop the asymptotic distribution of $\hat{\Pi}_{in}(m)$ and $\hat{S}_{in}(m)^{\dagger}$ for $i = 1, 2$.

Corollary 8.1. *Provided that* $\mathrm{E}[\|X\|_{L^2(E)}^4] < \infty$*, then for* $i = 1, 2$ *and* $m \geq 1$*,*

$$
\begin{aligned}
\sqrt{n} \quad (\hat{\Pi}_{in}(m) - \Pi_i(m)) &\xrightarrow{d} \mathcal{A}_i(m) \equiv \sum_{j>m} \sum_{\substack{k \neq j \\ k>m}} P_{ij}\mathcal{N}_i P_{ik} + \sum_{k>m} \sum_{\substack{j \neq k \\ j>m}} P_{ik}\mathcal{N}_i P_{ij} \\
= \quad & [I - \Pi_i(m)] \left(\sum_{j=1}^{\infty} (P_{ij}\mathcal{N}_i Q_{ij} + Q_{ij}\mathcal{N}_i P_{ij}) \right) [I - \Pi_i(m)] \qquad (70)
\end{aligned}
$$

where $\mathcal{N}_i$ *is a the distributional limit of* $\sqrt{n}(\hat{S}_{in} - S_i)$*.*

Proof: From Dauxois et al. [9] we know that for $i = 1, 2$

$$
\sqrt{n}\left\{\hat{P}_{ikn} - P_{ik}\right\} \xrightarrow{d} P_{ik}\mathcal{N}_i Q_{ik} + Q_{ik}\mathcal{N}_i P_{ik} \tag{71}
$$

where $\{\hat{P}_{ikn}, P_{ik}\}_{i=1}^{2}$, denote the eigenprojection operators for associated with the k^{th} largest eigenvalues to $\{\hat{S}_i, S_i\}_{i=1}^{2}$ and

$$Q_{ik} = \sum_{j \neq k} \frac{1}{\lambda_{ij} - \lambda_{ik}} P_{ij}.$$

For the asymptotic distribution of the cumulative eigenprojection operator $\hat{\Pi}_{in}(m) = \sum_{j=1}^{m} \hat{P}_{ijn}$, we notice that for all $m \geq 1$ the cumulative sum of the first term on the right hand side of (71) is

$$\sum_{j \leq m} P_{ij} \mathcal{N}_i Q_{1j} = \sum_{j \leq k} P_{ij} \mathcal{N}_i \sum_{k \neq j} \frac{1}{\lambda_{ik} - \lambda_{ij}} P_{ik}$$

and involves terms like

$$\begin{bmatrix} 0 & \frac{P_{i1}\mathcal{N}_i P_{i2}}{\lambda_{i2}-\lambda_{i1}} & \frac{P_{i1}\mathcal{N}_i P_{i3}}{\lambda_{i3}-\lambda_{i1}} & \cdots & \frac{P_{i1}\mathcal{N}_i P_{im}}{\lambda_{im}-\lambda_{i1}} & \cdots \\ \frac{P_{i2}\mathcal{N}_i P_{i1}}{\lambda_{i1}-\lambda_{i2}} & 0 & \frac{P_{i2}\mathcal{N}_i P_{i3}}{\lambda_{i3}-\lambda_{i2}} & \cdots & \frac{P_{i2}\mathcal{N}_i P_{1m}}{\lambda_{im}-\lambda_{i2}} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots \\ \frac{P_{im}\mathcal{N}_i P_{i1}}{\lambda_{i1}-\lambda_{im}} & \frac{P_{im}\mathcal{N}_i P_{i2}}{\lambda_{i2}-\lambda_{im}} & \frac{P_{im}\mathcal{N}_i P_{13}}{\lambda_{i3}-\lambda_{im}} & \cdots & 0 & \cdots \end{bmatrix}.$$

Hence for any term with $j, k \leq m$ and $j \neq k$ the upper triangular terms (UTT) involve $\frac{P_{ik}\mathcal{N}_i P_{ij}}{\lambda_{ij}-\lambda_{ik}}$ and the lower triangular terms (LTT) are $\frac{(P_{ij}\mathcal{N}_1 P_{ik})}{\lambda_{ik}-\lambda_{ij}} = \frac{-(P_{ik}\mathcal{N}_i P_{ij})^*}{\lambda_{ij}-\lambda_{ik}}$ so that $LTT = -UTT^*$. Since

$$\sqrt{n}(\hat{\Pi}_{in}(m) - \Pi_i(m)) \xrightarrow{d} \sum_{j \leq m} P_{ij} \mathcal{N}_i Q_{ij} + \sum_{j \leq m} Q_{ij} \mathcal{N}_i P_{ij} \tag{72}$$

and $(P_{ij}\mathcal{N}_i Q_{ij})^* = Q_{ij}\mathcal{N}_i P_{ij}$, the lower triangular terms in the first summand will cancel with the upper triangular terms in the second summand for all indices $i, j \leq m$. Equation (72) then telescopes and produces the following new asymptotic result

$$\begin{aligned} \sqrt{n} \quad & (\hat{\Pi}_i(m) - \Pi_i(m)) \xrightarrow{d} \sum_{j>m} \sum_{\substack{k \neq j \\ k>m}} P_{ij}\mathcal{N}_i P_{ik} + \sum_{j>m} \sum_{\substack{k \neq j \\ k>m}} P_{ik}\mathcal{N}_i P_{ij} \\ = \quad & [I - \Pi_i(m)] \left(\sum_{j=1}^{\infty} (P_{ij}\mathcal{N}_i Q_{ij} + Q_{ij}\mathcal{N}_i P_{ij}) \right) [I - \Pi_i(m)]. \end{aligned}$$

$\diamondsuit$

It is important to note that Corollary 8.1 has applications not just to canonical correlation analysis but also to principal component analysis.

Now, consider the asymptotic distribution of $\hat{S}_i(m)^\dagger = (\hat{S}_{in}\hat{\Pi}_i(m))^\dagger$ for some $m > 0$ and $i = 1, 2$. In this regard, observe that the function $F(z) = z^{-1}$ is analytic everywhere in the complex plane except for a pole at zero. Therefore F is analytic on the subset of the complex plane defined by $D_i \equiv \{z \in \mathbb{C} : \mathrm{Re}(z) \geq$

$\lambda_{im} - \epsilon\}$ with $0 < \epsilon < \lambda_{im}$. The set D_i also contains the spectrum of $S_i(m)$. Consequently, by the delta theorem (see Cupidon et al. [7] and appendix) we have that

$$\sqrt{n}\left\{\hat{S}_{in}(m)^{\dagger} - S_i(m)^{\dagger}\right\} \xrightarrow{d} \mathcal{B}_i(m) \tag{73}$$

for $i = 1, 2$ where

$$\mathcal{B}_i(m) \equiv -\sum_{j=1}^{m} \lambda_{ij}^{-2} P_{ij}\mathcal{N}_i P_{ij} + \sum_{\substack{k \neq j \\ j,k \leq m}} \frac{\lambda_{ik}^{-1} - \lambda_{ij}^{-1}}{\lambda_{ik} - \lambda_{ij}} P_{ik}\mathcal{N}_i P_{ij}. \tag{74}$$

The asymptotic analysis of $\sqrt{n}(\hat{\mathcal{S}}_{in}(m) - \mathcal{S}_i(m))$ for $i = 1, 2$ may now proceed where the application of the delta method leads to a product rule development. For this purpose we introduce the following Gaussian Hilbert-Schmidt operators

$$\begin{aligned}
\mathcal{F}_{11}(m) &\equiv \mathcal{A}_1(m)S_{12}S_2(m)^{\dagger}S_{21}S_1(m)^{\dagger}, \\
\mathcal{F}_{12}(m) &\equiv \Pi_1(m)\mathcal{N}_{12}S_2(m)^{\dagger}S_{21}S_1(m)^{\dagger}, \\
\mathcal{F}_{13}(m) &\equiv \Pi_1(m)S_{12}\mathcal{B}_2(m)S_{21}S_1(m)^{\dagger}, \\
\mathcal{F}_{14}(m) &\equiv \Pi_1(m)S_{12}S_2(m)^{\dagger}\mathcal{N}_{21}S_1(m)^{\dagger}, \\
\mathcal{F}_{15}(m) &\equiv \Pi_1(m)S_{12}S_2(m)^{\dagger}S_{21}\mathcal{B}_1(m), \\
\mathcal{F}_1(m) &\equiv \sum_{j=1}^{5} \mathcal{F}_{1j}(m) = \sum_{j=2}^{5} \mathcal{F}_{1j}(m).
\end{aligned} \tag{75}$$

The corollary below then results from the application of the delta theorem (see Cupidon et al. [7]).

Corollary 8.2. *If* $\mathrm{E}[\|X\|^4_{L^2(E)}] < \infty$*, then as* $n \to \infty$*,*

$$\sqrt{n}(\hat{\mathcal{S}}_1(m) - \mathcal{S}_1(m)) \xrightarrow{d} \mathcal{F}_1(m). \tag{76}$$

Proof: The proof follows along lines of the one for Corollary 6.1. Specifically, we begin by defining the elements

$$\begin{aligned}
\hat{\mathcal{A}}_{11}(m) &\equiv \left[\hat{\Pi}_{1n}(m) - \Pi_1(m)\right]\hat{S}_{12n}\hat{S}_{2n}(m)^{\dagger}\hat{S}_{21n}\hat{S}_{1n}(m)^{\dagger}, \\
\hat{\mathcal{A}}_{12}(m) &\equiv \Pi_1(m)\left[\hat{S}_{12n} - S_{12}\right]\hat{S}_{2n}(m)^{\dagger}\hat{S}_{21n}\hat{S}_{1n}(m)^{\dagger}, \\
\hat{\mathcal{A}}_{13}(m) &\equiv \Pi_1(m)S_{12}\left[\hat{S}_{2n}(m)^{\dagger} - S_2(m)^{\dagger}\right]\hat{S}_{21n}\hat{S}_{1n}(m)^{\dagger}, \\
\hat{\mathcal{A}}_{14}(m) &\equiv \Pi_1(m)S_{12}S_2(m)^{\dagger}\left[\hat{S}_{21n} - S_{21}\right]\hat{S}_{1n}(m)^{\dagger}, \\
\hat{\mathcal{A}}_{15}(m) &\equiv \Pi_1(m)S_{12}S_2(m)^{\dagger}S_{21}\left[\hat{S}_{1n}(m)^{\dagger} - S_1(m)^{\dagger}\right].
\end{aligned}$$

With this notation we may write

$$\sqrt{n}(\hat{\mathcal{S}}_{1n}(m) - \mathcal{S}_1(m)) = \sqrt{n}\left[\sum_{j=1}^{5} \hat{\mathcal{A}}_{1j}(m)\right].$$

The application of (44), (70), (73) and Slutsky's Theorem then ensure that

$$\sqrt{n}\left[\sum_{j=1}^{5}\hat{\mathcal{A}}_{1j}(m)\right] \xrightarrow{d} \mathcal{F}_1(m)$$

since, for example, the term $\hat{\mathcal{A}}_{11}(\alpha)$ consists of the factor $\sqrt{n}\left[\hat{\Pi}_{1n}(m) - \Pi_1(m)\right] \xrightarrow{d} \mathcal{A}_1(m)$ right-multiplied by the factor

$$\hat{S}_{12n}\hat{S}_{2n}(m)^{\dagger}\hat{S}_{21n}\hat{S}_{1n}(m)^{\dagger} \xrightarrow{p} S_{12}S_2(m)^{\dagger}S_{21}S_1(m)^{\dagger}.$$

We can now show that $\|\mathcal{F}_{11}(m)\| = 0$ with probability 1. To see this note that $\mathcal{F}_{11}(m)$ is self-adjoint because it is the distributional limit of self-adjoint operators. Furthermore, as a consequence of Corollary 8.1

$$\begin{aligned}
\mathcal{F}_{11}(m) &= \mathcal{A}_1(m)S_{12}S_2(m)^{\dagger}S_{21}S_1(m)^{\dagger} \\
&= [I - \Pi_1(m)]\,\mathcal{A}_1(m)S_{12}S_2(m)^{\dagger}S_{21}S_1(m)^{\dagger}\,[\Pi_1(m)] \\
&= [I - \Pi_1(m)]\,\mathcal{F}_{11}^{*}(m)\,[\Pi_1(m)] \\
&= [I - \Pi_1(m)]\,S_1(m)^{\dagger}S_{12}S_2(m)^{\dagger}S_{21}\mathcal{A}_1(m)\,[\Pi_1(m)] \\
&= [I - \Pi_1(m)]\,[\Pi_1(m)]\,S_1(m)^{\dagger}S_{12}S_2(m)^{\dagger}S_{21}\mathcal{A}_1(m)\,[I - \Pi_1(m)]\,[\Pi_1(m)] \\
&= 0.
\end{aligned}$$

Thus, $\|\mathcal{F}_{11}(m)\| = 0$ with probability 1 and $\mathcal{F}_1(m) = \sum_{j=2}^{5}\mathcal{F}_{1j}(m)$. This completes the proof. $\diamondsuit$

Corollary 8.2 ensures that for all $m \geq 1$

$$\|\hat{\mathcal{S}}_{1n}(m) - \mathcal{S}_1(m)\| = \mathcal{O}_P(n^{-1/2}) \xrightarrow{p} 0.$$

Hence, $\hat{\mathcal{S}}_{1n}(m)$ is consistent for $\mathcal{S}_1(m)$. The triangle inequality reveals the association between errors which originate from having a sample estimator and using regularization to approximate the desired operator TT^*,

$$\|\hat{\mathcal{S}}_{1n}(m) - TT^*\| \leq \|\hat{\mathcal{S}}_{1n}(m) - \mathcal{S}_1(m)\| + \|\mathcal{S}_1(m) - TT^*\|. \tag{77}$$

The first term on the right-hand side of (77) is a random error that originates from using a sample estimator of $\mathcal{S}_1(m)$ and tends to zero as $n \to \infty$. Meanwhile, the second term on the right hand side of (77) is a deterministic error that arises from using a regularized approximation of TT^* and will tend to zero as $m \uparrow \infty$.

Since the limiting distribution for $\sqrt{n}(\hat{\mathcal{S}}_{1n}(m) - \mathcal{S}_1(m))$ has been established, we may derive large sample asymptotics for $\{\hat{\rho}_{1jn}(m), \hat{\tilde{P}}_{1jn}(m), \hat{\tilde{f}}_{1jn}(m)\}$ where these quantities represent the j^{th} eigenvalue, eigenprojection and eigenvector for $\hat{\mathcal{S}}_{1n}(m)$. Let $\{\rho_{1j}(m), \tilde{P}_{1j}(m), \tilde{f}_{1j}(m)\}$ denote similar quantities for $\mathcal{S}_1(m)$.

We begin our development with the limiting distribution of the eigenprojection operators and associated eigenvectors.

Theorem 8.1. *Suppose* $\mathrm{E}\|X\|^4_{L^2(E)} < \infty$*. Then, for* $m \geq 1$ *and as* $n \to \infty$

$$\sqrt{n}\left\{\hat{\tilde{P}}_{1kn}(m) - \tilde{P}_{1k}(m)\right\} \xrightarrow{d} \tilde{P}_{1k}(m)\mathcal{F}_1(m)\tilde{Q}_{1k}(m) + \tilde{Q}_{1k}(m)\mathcal{F}_1(m)\tilde{P}_{1k}(m) \tag{78}$$

where $\mathcal{F}_1(m)$ *is as in (76) and*

$$\tilde{Q}_{1k}(m) = \sum_{j\neq k} \frac{1}{\rho_{1j}(m) - \rho_{1k}(m)}\tilde{P}_{1k}(m).$$

In the case that $\mathrm{rank}(\tilde{P}_{1k}(m)) = 1$*, then*

$$\sqrt{n}\left\{\hat{\tilde{f}}_{1kn}(m) - \tilde{f}_{1k}(m)\right\} \xrightarrow{d} \tilde{Q}_{1k}(m)\mathcal{F}_1(m)\tilde{f}_{1k}(m). \tag{79}$$

Proof: The proof for the limiting distribution of $\hat{\tilde{P}}_{1kn}(m)$ is identical to that presented for Tikhinov regularization in Theorem 6.1. The only difference is that the role of the parameter α in Tikhinov regularization is replaced by that of m in TSVD regularization. For the sake of completeness, we provide a sketch of the proof.

For each $k \geq 1$, let Γ_k denote a circle that encloses the eigenvalue $\rho_{1k}(m)$ but no other eigenvalues of $\mathcal{S}_1(m)$. From developments in the appendix, notice that

$$\sqrt{n}\left\{\hat{\tilde{P}}_{1kn}(m) - \tilde{P}_{1k}(m)\right\} = \frac{\sqrt{n}}{2\pi i}\oint_{\Gamma_j} R(z)(\hat{\mathcal{S}}_{1n}(m) - \mathcal{S}_1(m))R(z)dz + \mathcal{O}_P(n^{-1/2}). \tag{80}$$

Focussing attention on the first term on the right hand side of (80), it follows from the continuous mapping theorem that

$$\sqrt{n}\left\{\hat{\tilde{P}}_{1kn}(m) - \tilde{P}_{1k}(m)\right\} \xrightarrow{d} \frac{1}{2\pi i}\oint_{\Gamma_k} R(z)\mathcal{F}_1(m)R(z)dz. \tag{81}$$

Since

$$R(z) = \sum_{k=1}^{\infty} \frac{1}{\rho_k(m) - z}\tilde{P}_{1k}(m) + \mathcal{O}((\rho_k(m) - z)^{-2}) \tag{82}$$

all but the lead term in (82) will vanish when the contour integral is taken due to (96). The integrand in (81) can then be simplified as

$$R(z)\mathcal{F}_1(m)R(z) = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \frac{1}{(\rho_k(m) - z)(\rho_j(m) - z)}\tilde{P}_{1k}(m)\mathcal{F}_1(m)\tilde{P}_{1j}(m).$$

Using the Cauchy integral formula produces

$$\frac{1}{2\pi i}\oint_{\Gamma_k} \frac{dz}{(\rho_k(m) - z)(\rho_j(m) - z)} = \frac{1}{2\pi i}\oint_{\Gamma_k} \frac{(\rho_i(m) - z)dz}{(\rho_k(m) - z)(\rho_j(m) - z)(\rho_i(m) - z)},$$

and the only case where the integral is non-zero is when exactly one of $\rho_k(m)$ or $\rho_j(m)$ is not equal to $\rho_i(m)$. Hence

$$\begin{aligned}\sqrt{n}\left\{\hat{\tilde{P}}_{1kn}(m)-\tilde{P}_{1k}(m)\right\} &\xrightarrow{d} \sum_{i=1}^{\infty}\sum_{j\neq k}\frac{\delta_{ik}}{(\rho_j(m)-\rho_k(m))}\tilde{P}_{1i}(m)\mathcal{F}_1(m)\tilde{P}_{1j}(m)\\ &+ \sum_{j=1}^{\infty}\sum_{i\neq k}\frac{\delta_{jk}}{(\rho_i(m)-\rho_k(m))}\tilde{P}_{1i}(m)\mathcal{F}_1(m)\tilde{P}_{1j}(m)\\ &= \tilde{P}_{1k}(m)\mathcal{F}_1(m)\tilde{Q}_{1k}(m)+\tilde{Q}_{1k}(m)\mathcal{F}_1(m)\tilde{P}_{1k}(m)\end{aligned}$$

which establishes (78).

To establish the limiting distribution of the eigenvectors in (79) we write

$$\begin{aligned}\sqrt{n}\left\{\hat{\tilde{f}}_{1kn}(m)-\tilde{f}_{1k}(m)\right\} &= \sqrt{n}\left[\tilde{P}_{1k}(m)\right]\left\{\hat{\tilde{f}}_{1kn}(m)-\tilde{f}_{1k}(m)\right\}\\ &+ \sqrt{n}\left[I-\tilde{P}_{1k}(m)\right]\left\{\hat{\tilde{f}}_{1kn}(m)-\tilde{f}_{1k}(m)\right\}. \quad (83)\end{aligned}$$

Now by using the TSVD analogues to equations (60) and (61) we may see that the limiting distribution for the first term on the right hand side of (83) is 0. For the second term on the right hand side of (83) we have

$$\sqrt{n}\left[I-\tilde{P}_{1k}(m)\right]\left\{\hat{\tilde{f}}_{1kn}(m)-\tilde{f}_{1k}(m)\right\}\xrightarrow{d}\tilde{Q}_{1j}(m)\mathcal{F}_1(m)\tilde{f}_{1k}(m),$$

which completes the proof. $\diamondsuit$

We will now derive the limiting distribution for $\sqrt{n}\,[\hat{\rho}_{1kn}(m)-\rho_{1k}(m)]$, where $\{\hat{\rho}_{1kn}(m),\rho_{1k}(m)\}$ denotes the k^{th} distinct eigenvalue associated with $\{\hat{\mathcal{S}}_{1n}(m),\mathcal{S}_1(m)\}$. Much like the Tikhinov case, the quantity $\sqrt{n}\,[\hat{\rho}_{1kn}(m)-\rho_{1k}(m)]$ will be regarded as a vector of dimension equal to the multiplicity, d_k, of the eigenvalue $\rho_{1k}(m)$.

Theorem 8.2. *Assume that* $\mathrm{E}\|X\|^4_{L^2(E)}<\infty$ *and the* k^{th} *regularized canonical correlation,* $\rho_{1k}(m)$*, has geometric multiplicity* d_k*. Then*

$$\sqrt{n}\,[\hat{\rho}_{1kn}(m)-\rho_{1k}(m)]\xrightarrow{d}\tilde{P}_{1k}(m)\mathcal{F}_1(m)\tilde{P}_{1k}(m) \quad (84)$$

with $\mathcal{F}_1(m)$ *the Gaussian random variable in (75). Furthermore,* $\tilde{P}_{1k}(m)\mathcal{F}_1(m)\tilde{P}_{1k}(m)$ *has dimension* d_k*. In the special case that* $d_k=1$

$$\sqrt{n}\,(\hat{\rho}_{1kn}(m)-\rho_{1k}(m))\xrightarrow{d}N(0,\sigma_{kk}(m)) \quad (85)$$

where $N(0,\sigma_{kk}(m))$ *denotes a normal distribution with zero mean and variance*

$$\sigma_{kk}(m)=\mathrm{E}\left[\langle\tilde{f}_{1k}(m),\mathcal{F}_1(m)\tilde{f}_{1k}(m)\rangle^2_{\mathcal{H}(K_1)}\right].$$

Proof: Like before, the proof here naturally parallels the Tikhinov result presented in Theorem 6.2. Since $\|\hat{\tilde{P}}_{1kn}(m) - \tilde{P}_{1k}(m)\| = \mathcal{O}_P(n^{-1/2})$, Theorem 10.1 ensures that for large enough n, $\text{rank}(\hat{\tilde{P}}_{1kn}(m)) = \text{rank}(\tilde{P}_{1k}(m)) = d_k$ with probability tending to 1 as $n \to \infty$. Now let us define

$$\begin{aligned}
\hat{\mathcal{D}}_{k1}(m) &\equiv [\hat{\tilde{P}}_{1kn}(m) - \tilde{P}_{1k}(m)]\hat{\mathcal{S}}_{1n}(m)\hat{\tilde{P}}_{1kn}(m), \\
\hat{\mathcal{D}}_{k2}(m) &\equiv \tilde{P}_{1k}(m)[\hat{\mathcal{S}}_{1n}(m) - \mathcal{S}_1(m)]\hat{\tilde{P}}_{1kn}(m), \\
\hat{\mathcal{D}}_{k3}(m) &\equiv \tilde{P}_{1k}(m)\mathcal{S}_1(m)[\hat{\tilde{P}}_{1kn}(m) - \tilde{P}_{1k}(m)],
\end{aligned}$$

and note that

$$\sqrt{n}\,[\hat{\rho}_{1kn}(m) - \rho_{1k}(m)] = \sqrt{n}\left[\sum_{j=1}^{3} \hat{\mathcal{D}}_{kj}(m)\right]. \tag{86}$$

Note that $\hat{\mathcal{D}}_{k1}(m) \xrightarrow{p} 0$ and $\hat{\mathcal{D}}_{k3}(m) \xrightarrow{p} 0$ since

$$\begin{aligned}
& \|\hat{\mathcal{D}}_{k1}(m)\|_{HS}^2 \xrightarrow{d} \| \left[\tilde{Q}_{1k}(m)\mathcal{F}_1(m)\tilde{P}_{1k}(m) + \tilde{P}_{1k}(m)\mathcal{F}_1(m)\tilde{Q}_{1k}(m)\right] \mathcal{S}_1(m)\tilde{P}_{1k}(m)\|_{HS}^2 \\
\leq \;& \|\tilde{Q}_{1k}(m)\mathcal{F}_1(m)\mathcal{S}_1(m)\tilde{P}_{1k}(m)\|_{HS}^2 + \|\tilde{P}_{1k}(m)\mathcal{F}_1(m)\tilde{Q}_{1k}(m)\mathcal{S}_1(m)\tilde{P}_{1k}(m)\|_{HS}^2 \\
= \;& \|\tilde{Q}_{1k}(m)\mathcal{F}_1(m)\mathcal{S}_1(m)\tilde{P}_{1k}(m)\|_{HS}^2 + \|\tilde{P}_{1k}(m)\mathcal{F}_1(m)\left[\tilde{Q}_{1k}(m)\tilde{P}_{1k}(m)\right]\mathcal{S}_1(m)\|_{HS}^2 \\
= \;& \text{tr}\left(\tilde{P}_{1k}(m)\mathcal{S}_1(m)\mathcal{F}_1(m)\tilde{Q}_{1k}^2(m)\mathcal{F}_1(m)\mathcal{S}_1(m)\tilde{P}_{1k}(m)\right) + 0 \\
= \;& \text{tr}\left(\mathcal{S}_1(m)\left[\tilde{P}_{1k}(m)\tilde{Q}_{1k}(m)\right]\mathcal{F}_1(m)\mathcal{F}_1(m)\left[\tilde{Q}_{1k}(m)\tilde{P}_{1k}(m)\right]\mathcal{S}_1(m)\right) = 0.
\end{aligned}$$

Similarly,

$$\begin{aligned}
& \|\hat{\mathcal{D}}_{k3}(m)\|_{HS}^2 \xrightarrow{d} \|\tilde{P}_{1k}(m)\mathcal{S}_1(m)\left[\tilde{Q}_{1k}(m)\mathcal{F}_1(m)\tilde{P}_{1k}(m) + \tilde{P}_{1k}(m)\mathcal{F}_1(m)\tilde{Q}_{1k}(m)\right]\|_{HS}^2 \\
\leq \;& \|\mathcal{S}_1(m)\left[\tilde{P}_{1k}(m)\tilde{Q}_{1k}(m)\right]\mathcal{F}_1(m)\tilde{P}_{1k}(m)\|_{HS}^2 + \|\tilde{P}_{1k}(m)\mathcal{S}_1(m)\mathcal{F}_1(m)\tilde{Q}_{1k}(m)\|_{HS}^2 \\
= \;& 0 + \text{tr}\left(\tilde{Q}_{1k}(m)\mathcal{F}_1(m)\mathcal{S}_1(m)\tilde{P}_{1k}(m)\mathcal{S}_1(m)\mathcal{F}_1(m)\tilde{Q}_{1k}(m)\right) \\
= \;& \text{tr}\left(\mathcal{F}_1(m)\left[\tilde{Q}_{1k}(m)\tilde{P}_{1k}(m)\right]\mathcal{S}_1(m)\tilde{P}_{1k}(m)\mathcal{S}_1(m)\left[\tilde{P}_{1k}(m)\tilde{Q}_{1k}(m)\right]\mathcal{F}_1(m)\right) = 0.
\end{aligned}$$

Hence, Corollary 8.2 and Slutsky's Theorem ensure that

$$\sqrt{n}\hat{\mathcal{D}}_{k2}(m) \xrightarrow{d} \tilde{P}_{1k}(m)\mathcal{F}_1(m)\tilde{P}_{1k}(m)$$

which proves the first part of the theorem.

To see the validity of the second part of the theorem, assume that $d_k = 1$ and observe that

$$\sqrt{n}\,(\hat{\rho}_{kn}(m) - \rho_k(m)) = \sqrt{n}\left\{\sum_{j=1}^{3} \hat{\mathcal{C}}_{kj}(m)\right\}$$

where

$$\hat{\mathcal{C}}_{k1}(m) \equiv \langle [\hat{\tilde{f}}_{kn}(m) - \tilde{f}_k(m)], \hat{\mathcal{S}}_{1n}(m)\hat{\tilde{f}}_{kn}(m)\rangle_{\mathcal{H}(K_1)},$$
$$\hat{\mathcal{C}}_{k2}(m) \equiv \langle \tilde{f}_k(m), [\hat{\mathcal{S}}_{1n}(m) - \mathcal{S}_1(m)]\hat{\tilde{f}}_{kn}(m)\rangle_{\mathcal{H}(K_1)},$$
$$\hat{\mathcal{C}}_{k3}(m) \equiv \langle \tilde{f}_k(m), \mathcal{S}_1(m)[\hat{\tilde{f}}_{kn}(m) - \tilde{f}_k(m)]\rangle_{\mathcal{H}(K_1)}.$$

The terms $\hat{\mathcal{C}}_{k1}(m) \xrightarrow{d} 0$ and $\hat{\mathcal{C}}_{k3}(m) \xrightarrow{d} 0$ as a consequence of equation (79) and Slutsky's theorem, since

$$\begin{aligned}
\hat{\mathcal{C}}_{k1}(m) &\xrightarrow{d} \langle \tilde{Q}_{1k}(m)\mathcal{F}_1(m)\tilde{f}_k(m), \mathcal{S}_1(m)\tilde{f}_k(m)\rangle_{\mathcal{H}(K_1)} \\
&= \sum_{j\neq k} \frac{\rho_k(m)}{(\rho_j(m)-\rho_k(m))}\langle \tilde{P}_{1j}(m)\mathcal{F}_1(m)\tilde{f}_k(m), \tilde{f}_k(m)\rangle_{\mathcal{H}(K_1)} = 0
\end{aligned}$$

and

$$\begin{aligned}
\hat{\mathcal{C}}_{k3}(m) &\xrightarrow{d} \langle \tilde{f}_k(m), \mathcal{S}_1(m)\tilde{Q}_{1k}(m)\mathcal{F}_1(m)\tilde{f}_k(m)\rangle_{\mathcal{H}(K_1)} \\
&= \sum_{j\neq k} \frac{\rho_k(m)}{(\rho_j(m)-\rho_k(m))}\langle \tilde{P}_{1j}(m)\tilde{f}_k(m), \mathcal{F}_1(m)\tilde{f}_k(m)\rangle_{\mathcal{H}(K_1)} = 0.
\end{aligned}$$

Application of Theorem 8.2 implies that

$$\sqrt{n}\hat{\mathcal{C}}_{k2}(m) \xrightarrow{d} \langle \tilde{f}_k(m), \mathcal{F}_1(m)\tilde{f}_k(m)\rangle_{\mathcal{H}(K_1)}.$$

Since

$$\mathrm{E}\left[\langle \tilde{f}_k(m), \mathcal{F}_1(m)\tilde{f}_k(m)\rangle_{\mathcal{H}(K_1)}\right] = 0$$

and

$$\mathrm{Var}\left[\langle \tilde{f}_k(m), \mathcal{F}_1(m)\tilde{f}_k(m)\rangle_{\mathcal{H}(K_1)}\right] = \mathrm{E}\left[\langle \tilde{f}_k(m), \mathcal{F}_1(m)\tilde{f}_k(m)\rangle^2_{\mathcal{H}(K_1)}\right] \equiv \sigma_{kk}(m),$$

the proof is then complete. $\diamondsuit$

The TSVD versions of the correlation estimators $\{\hat{\rho}_{in}(m), \hat{\rho}_{jn}(m)\}$ with $i \neq j$ are correlated, much like the Tikhinov case. In fact, when the operator $\hat{\mathcal{S}}_{1n}(m)$ is simple, we have for $i \neq j$

$$\begin{aligned}
&\sigma_{ij}(m) \equiv \mathrm{Cov}[\hat{\rho}_{in}(m), \hat{\rho}_{jn}(m)] \\
=\ & \mathrm{E}\left[(\hat{\tilde{f}}_{in}(m), \hat{\mathcal{S}}_{1n}(m)\hat{\tilde{f}}_{in}(m))_{\mathcal{H}(K_1)}, (\hat{\tilde{f}}_{jn}(m), \hat{\mathcal{S}}_{1n}(m)\hat{\tilde{f}}_{jn}(m))_{L^2(E_1)}\right] \\
=\ & \mathrm{E}\left[\left(\left(\hat{\tilde{f}}_{in}(m) \otimes_{\mathcal{H}(K_1)} \hat{\tilde{f}}_{in}(m)\right), \left[\hat{\mathcal{S}}_{1n}(m) \otimes_{HS_1} \hat{\mathcal{S}}_{1n}(m)\right]\left(\hat{\tilde{f}}_{jn}(m) \otimes_{\mathcal{H}(K_1)} \hat{\tilde{f}}_{jn}(m)\right)\right)_{HS_1}\right] \\
=\ & \left(\left(\tilde{f}_i(m) \otimes_{\mathcal{H}(K_1)} \tilde{f}_i(m)\right), \mathrm{E}\left[\hat{\mathcal{S}}_{1n}(m) \otimes_{HS_1} \hat{\mathcal{S}}_{1n}(m)\right]\left(\tilde{f}_j(m) \otimes_{\mathcal{H}(K_1)} \tilde{f}_j(m)\right)\right)_{HS_1} \\
=\ & \left(\left(\tilde{f}_i(m) \otimes_{\mathcal{H}(K_1)} \tilde{f}_i(m)\right), \Sigma_1(m)\left(\tilde{f}_j(m) \otimes_{\mathcal{H}(K_1)} \tilde{f}_j(m)\right)\right)_{HS_1} \\
\equiv\ & [\Sigma_1(m)]_{ij}
\end{aligned}$$

with

$$\Sigma_1(m) \equiv \mathrm{E}\left[\hat{\mathcal{S}}_{1n}(m) \otimes_{HS_1} \hat{\mathcal{S}}_{1n}(m)\right]$$

and $[\Sigma_1(m)]_{ij}$, the $\{i,j\}^{th}$ element of $\Sigma_1(m)$.

9. Conclusion

In Sections 5–9 we discussed how both Tikhinov and TSVD regularized estimators approach their intended target of approximation, the RKHS based operator TT^*, in the limits of their respective regularization parameters. We also showed that the asymptotics associated with Tikhinov and TSVD sample estimators $\{\hat{\mathcal{S}}_{1n}(\alpha), \hat{\mathcal{S}}_{1n}(m)\}$ are similar in the sense that for every distributional result for quantities relative to the Tikhinov estimator $\hat{\mathcal{S}}_{1n}(\alpha)$, there is an analogous distributional result for its TSVD cousin $\hat{\mathcal{S}}_{1n}(m)$. The question to ask here is whether or not one form of regularization should be preferred over the other.

The answer to this question lies in one critical flaw in the Tikhinov approach to FCCA, which up to this point has not yet been discussed. Although replacing the operators $\{S_1, S_2\}$ with $\{(S_1+\alpha I), (S_2+\alpha I)\}$ fixes the operators invertibility issues, the operators still theoretically have infinite dimensionality. Infinite dimensional operators are problematic because no computer will ever be able to estimate all the eigenvalues and eigenvectors. On the sample side, since the operators $\{\hat{S}_{1n}, \hat{S}_{2n}\}$ have rank at most n, they are rank deficient. Meanwhile the operators $\{(\hat{S}_{1n}+\alpha I), (\hat{S}_{2n}+\alpha I)\}$ will have infinitely many eigenvalues equal to α. Any pragmatic computational scheme where Tikhinov regularization is implemented would therefore involve some limit on the number of eigenvalue and eigenvector pairs to be used and estimated. As a consequence, FCCA methods will surely involve truncation. If we choose to implement Tikhinov regularization with truncation this will involve the operator

$$\hat{S}_{1n}(\alpha, m) \equiv \sum_{j=1}^{m} (\hat{\lambda}_{1jn} + \alpha)\hat{P}_{1jn} = (\hat{S}_{1n} + \alpha I)\hat{\Pi}_{1n}(m) \tag{87}$$

for some integer $1 \leq m \leq n$. The estimator in (87) has some characteristics that are akin to both those of Tikhinov and TSVD regularization. Utilizing this "truncated Tikhinov" estimator it follows that the corresponding regularized estimator for TT^* would be

$$\hat{\mathcal{S}}_{1n}(\alpha, m) \equiv \hat{\Pi}_{1n}(m)\hat{S}_{12n}\hat{S}^{\dagger}_{2n}(\alpha, m)\hat{S}_{21n}\hat{S}^{\dagger}_{1n}(\alpha, m). \tag{88}$$

Equation (88) illustrates that pragmatic implementation of Tikhinov regularization in the FDA setting will in reality entail the use of both Tikhinov and TSVD forms of regularization. By contrast, TSVD regularization entails replacing the operators $\{S_1, S_2\}$ with $\{(S_1\Pi_1(m), S_2\Pi_2(m)\}$ which have finite rank. Consequently, TSVD regularization provides a remedy for both infinite dimensionality and invertibility issues simultaneously.

Since there are errors which originate from regularization methods in general, it is always better to use as few methods as possible. The triangle inequality can now be utilized to establish a bound on the error associated with the "truncated Tikhinov" estimator (87). In this regard, notice that

$$\|\hat{\mathcal{S}}_{1n}(\alpha,m)-TT^*\|\leq\|\hat{\mathcal{S}}_{1n}(\alpha,m)-\hat{\mathcal{S}}_{1n}(m)\|+\|\hat{\mathcal{S}}_{1n}(m)-TT^*\|.$$

Hence the error associated with utilizing $\hat{S}_{1n}(\alpha,m)$ will always be larger than simply using $\hat{S}_{1n}(m)$.

10. Appendix: Some Perturbation Theory

In this appendix we briefly summarize some results from perturbation theory. The primary references for this section are Kato [22] and Dauxois et al. [9]. A typical problem in perturbation theory is to determine how the eigenvalues and eigenspaces of a linear operator B change when B is subjected to a small perturbation. Let $A:\mathcal{H}\mapsto\mathcal{H}$ be an arbitrary perturbation operator and let $\tilde{B}=B+A$ represent the perturbed operator. In this regard, we might think of A as being small in terms of its uniform operator norm $\|A\|$. However, a measure of "closeness" between $\tilde{B}$ and B which is often of greater importance is the aperture or gap between the graphs of the two operators.

Let $\mathcal{M}$ and $\mathcal{N}$ be two closed linear manifolds on $\mathcal{H}$ with $S_{\mathcal{M}}=\{u\in\mathcal{M}\mid\|u\|_{\mathcal{H}}=1\}$, the unit sphere on $\mathcal{M}$. For any two closed linear manifolds $\mathcal{M},\mathcal{N}\subset\mathcal{H}$ let

$$\delta(\mathcal{M},\mathcal{N})\equiv\begin{cases}\sup_{u\in S_{\mathcal{M}}}\{\mathrm{dist}(u,\mathcal{N})\} & \text{for } \mathcal{M}\neq\{0\},\\ 0 & \text{if } \mathcal{M}=\{0\}\end{cases}$$

with

$$\mathrm{dist}(u,\mathcal{N})\equiv\inf_{v\in\mathcal{N}}\{\|u-v\|_{\mathcal{H}}\}.$$

The gap between $\mathcal{M}$ and $\mathcal{N}$ is then defined by

$$\hat{\delta}(\mathcal{M},\mathcal{N})\equiv\max[\delta(\mathcal{M},\mathcal{N}),\delta(\mathcal{N},\mathcal{M})].$$

More details concerning $\delta(\mathcal{M},\mathcal{N})$ and $\hat{\delta}(\mathcal{M},\mathcal{N})$ can be found in Kato [22].

If the graphs $\{G(B),G(\tilde{B})\}$ of two operators $\{B,\tilde{B}\}$ are closed, the closed graph theorem entails that both B and $\tilde{B}$ are bounded. Consequently it is possible to define the gap between operators B and $\tilde{B}$ by measuring the gap between their associated graphs. In this regard we define

$$\delta(B,\tilde{B})\equiv\delta(G(B),G(\tilde{B})),$$

$$\hat{\delta}(B,\tilde{B})\equiv\hat{\delta}(G(B),G(\tilde{B}))=\max[\delta(B,\tilde{B}),\delta(\tilde{B},B)],$$

and $\hat{\delta}(B,\tilde{B})=\hat{\delta}(\tilde{B},B)$ is called the gap between B and $\tilde{B}$.

The notion of the gap between operators plays a large role in perturbation theory. Suppose B and $\tilde{B}$ are the original and perturbed operator respectively.

The smaller the gap $\hat{\delta}(\tilde{B}, B)$ becomes, the more properties the $\tilde{B}$ inherits from B. Of particular importance is the following theorem from Kato [22] which permits the construction of closed curve Γ around a part of the spectrum of B, denoted $\Sigma(B)$, that also encloses a similar collection of spectral points of the perturbed operator $\Sigma(\tilde{B})$.

Theorem 10.1. *(Semi-continuity of the spectrum) Let $\tilde{B}, B \in \mathcal{B}(\mathcal{H})$ and let the spectrum of B, $\Sigma(B)$, be separated into two parts $\Sigma'(B)$, $\Sigma''(B)$ by a closed curve Γ, with $\mathcal{H} = \mathcal{M}'(B) \oplus \mathcal{M}''(B)$. Then, there exists a $\delta > 0$, depending on Γ and B, such that if $\tilde{B}$ is any operator with $\hat{\delta}(\tilde{B}, B) < \delta$*

(i) *the spectrum $\Sigma(\tilde{B})$ are likewise separated by Γ into two parts $\{\Sigma'(\tilde{B}), \Sigma''(\tilde{B})\}$ and both $\{\Sigma'(\tilde{B}), \Sigma''(\tilde{B})\}$ are non-empty if this is true for $\{\Sigma'(B), \Sigma''(B)\}$,*
(ii) *in the associated decomposition $\mathcal{H} = \mathcal{M}'(\tilde{B}) \oplus \mathcal{M}''(\tilde{B})$, $\{\mathcal{M}'(\tilde{B}), \mathcal{M}''(\tilde{B})\}$ are isomorphic with $\{\mathcal{M}'(B), \mathcal{M}''(B)\}$, respectively,*
(iii) $\dim(\mathcal{M}'(\tilde{B})) = \dim(\mathcal{M}'(B))$ *and* $\dim(\mathcal{M}''(\tilde{B})) = \dim(\mathcal{M}''(B))$ *and*
(iv) *the projection operator $P_{\tilde{B}}$ of $\mathcal{H}$ onto $\mathcal{M}'(\tilde{B})$ tends to the similarly defined projection operator P_B in operator norm as $\hat{\delta}(\tilde{B}, B) \to 0$.*

We will now develop formulae for the differences in the resolvents and projection operators between the perturbed and unperturbed operator. In this regard, let $R(z) = (B - zI)^{-1}$ and $\tilde{R}(z') = (\tilde{B} - z'I)^{-1}$ denote the resolvents of B and $\tilde{B}$ for some $z \in \mathbb{C} \setminus \Sigma(B)$ and $z' \in \mathbb{C} \setminus \Sigma(\tilde{B})$, respectively. From Kato [22], if $\lambda_k \in \Sigma(B)$ is some isolated point of the spectra and P_k is the associated projection operator then

$$P_k = -\frac{1}{2\pi i} \oint_{\Gamma_k} R(z) dz \tag{89}$$

where Γ_k is a positively oriented curve that encloses λ_k but no other spectral values of $\Sigma(B)$.

Now, whenever $\|(\tilde{B} - B)R(z)\| < 1$ and $z \in \mathbb{C} \setminus \Sigma(B)$ we may utilize the Neumann series Theorem (Rynne and Youngson [33]) which ensures that

$$\begin{aligned}
\tilde{R}(z) &= \Big((\tilde{B} - B) + (B - zI)\Big)^{-1} \\
&= \Big((\tilde{B} - B) + (R(z))^{-1}\Big)^{-1} \\
&= R(z)\Big((\tilde{B} - B)R(z) + I\Big)^{-1} \\
&= R(z)\left(I + \sum_{k=1}^{\infty}\Big\{(B - \tilde{B})R(z)\Big\}^k\right).
\end{aligned}$$

It then follows that

$$\begin{aligned}
\Big[\tilde{R}(z) - R(z)\Big] &= R(z)(B - \tilde{B})R(z)\left[\sum_{k=0}^{\infty}\Big\{(B - \tilde{B})R(z)\Big\}^k\right] \\
&= R(z)(B - \tilde{B})R(z)H(z) \qquad (90)
\end{aligned}$$

where $H(z) \equiv \sum_{k=0}^{\infty} \left\{(B - \tilde{B})R(z)\right\}^k$. Another application of Neumann series theorem reveals that

$$H(z) = \sum_{k=0}^{\infty} \left\{(B - \tilde{B})R(z)\right\}^k = \left[I - (B - \tilde{B})R(z)\right]^{-1}$$

and hence

$$\left[\tilde{R}(z) - R(z)\right] = R(z)(B - \tilde{B})R(z)\left[I - (B - \tilde{B})R(z)\right]^{-1}.$$

Now let $\{\lambda_j, \tilde{\lambda}_j\}$ be particular spectral values for $\{B, \tilde{B}\}$, and let $\{P_j, \tilde{P}_j\}$ denote the corresponding eigenprojection operators. Now provided $\hat{\delta}(\tilde{B}, B)$ is small enough, Theorem 10.1 ensures that a positively oriented circle Γ_j, with radius r, can be drawn to enclose both λ_j and $\tilde{\lambda}_j$ but no other spectral values of either B or $\tilde{B}$. As a consequence of (89) and (90) we then obtain

$$\begin{aligned} \left[\tilde{P}_j - P_j\right] &= -\frac{1}{2\pi i}\oint_{\Gamma_j}\left[\tilde{R}(z) - R(z)\right]dz \\ &= -\frac{1}{2\pi i}\oint_{\Gamma_j} R(z)(B - \tilde{B})R(z)H(z)dz. \end{aligned} \tag{91}$$

Equation (91) allows us to formulate a crude bound on the uniform operator norm of $\|\tilde{P}_j - P_j\|$, specifically

$$\begin{aligned} \|\tilde{P}_j - P_j\| &\le \frac{1}{2\pi}\oint_{\Gamma_j} \|R(z)(B - \tilde{B})R(z)\left[I - (B - \tilde{B})R(z)\right]^{-1}\|dz \\ &\le \frac{2\pi r}{2\pi}\sup\left\{\frac{\|B - \tilde{B}\|\|R(z)\|^2}{1 - \|B - \tilde{B}\|\|R(z)\|} : z \in \Gamma_j\right\} \\ &= r\sup\left\{\frac{\|B - \tilde{B}\|\|R(z)\|^2}{1 - \|B - \tilde{B}\|\|R(z)\|} : z \in \Gamma_j\right\}. \end{aligned} \tag{92}$$

Another formula for $\left[\tilde{P}_j - P_j\right]$ can be derived by expanding the first term of $H(z)$ so that

$$H(z) = I + \sum_{j=1}^{\infty}\left\{(B - \tilde{B})R(z)\right\}^j. \tag{93}$$

Plugging (93) into (91) gives

$$\left[\tilde{P}_j - P_j\right] = -\frac{1}{2\pi i}\oint_{\Gamma_j} R(z)(B - \tilde{B})R(z)dz - \frac{1}{2\pi i}\oint_{\Gamma_j} M(z)dz \tag{94}$$

where

$$M(z) \equiv R(z)\sum_{j=2}^{\infty}\{-AR(z)\}^j = \mathcal{O}(A^2).$$

Using the partial fraction expansion of the resolvent from Kato [22] it follows that

$$R(z) = \sum_{j=1}^{\infty} \frac{1}{(\lambda_j - z)} P_j + \mathcal{O}((\lambda_j - z)^{-2}). \tag{95}$$

Now in (95) the higher-order terms involving $\mathcal{O}((\lambda_j - z)^{-2})$ can be ignored due to Morera's theorem since, for $n \geq 2$,

$$\oint_{\Gamma_j} (\lambda_j - z)^{-n} dz = \oint_{\Gamma_j} w^{n-2} dw = 0 \tag{96}$$

where the substitution $w = (\lambda_j - z)^{-1}$ has been used. Thus, since

$$\frac{1}{2\pi i} \oint_{\Gamma_j} \frac{1}{\lambda_j - z} \frac{1}{\lambda_k - z} dz = \begin{cases} \frac{1}{\lambda_j - \lambda_k} & \text{if } k \neq j, \\ 0 & \text{otherwise} \end{cases}$$

it follows that

$$\begin{aligned} \frac{1}{2\pi i} \oint_{\Gamma_j} R(z) A R(z) dz &= \frac{1}{2\pi i} \sum_k \sum_j \oint_{\Gamma_j} \frac{1}{\lambda_k - z} \frac{1}{\lambda_j - z} dz P_k A P_j \\ &= \sum_{k \neq l} \frac{1}{\lambda_j - \lambda_k} (P_k A P_j + P_j A P_k). \end{aligned}$$

Therefore

$$\tilde{P}_j - P_j = \sum_{k \neq j} \frac{1}{\lambda_j - \lambda_k} (P_k A P_j + P_j A P_k) + \frac{1}{2\pi i} \oint_{\Gamma} \phi(z) M(z) dz. \tag{97}$$

Equation (97) has several important implications as it allows us to formulate the notion of the Frechet derivative of an analytic function of an operator. Now suppose that a function $\phi(z)$ is analytic in a domain Δ of the complex plane containing all the spectral values $\{\lambda_h, \tilde{\lambda}_h\}$ of $\{B, \tilde{B}\}$, with $\Gamma \subset \Delta$ a positively oriented closed curve that encloses all spectral values in its interior. Utilizing the Dunsford-Taylor integral for $\phi(\tilde{B})$ and $\phi(B)$ (see Kato [22]) we obtain

$$\begin{aligned} \phi(\tilde{B}) - \phi(B) &= -\frac{1}{2\pi i} \oint_{\Gamma} \phi(z) [\tilde{R}(z) - R(z)] dz \\ &= \frac{1}{2\pi i} \oint_{\Gamma} \phi(z) R(z) A R(z) dz + \frac{1}{2\pi i} \oint_{\Gamma} \phi(z) M(z) dz \\ &= \frac{1}{2\pi i} \sum_k \sum_j \oint_{\Gamma} \frac{\phi(z)}{(\lambda_k - z)(\lambda_j - z)} dz P_k A P_j + \mathcal{O}(A^2). \end{aligned} \tag{98}$$

Focussing on the integral in the first term on the right hand side we see that

$$\frac{1}{2\pi i} \oint_{\Gamma} \frac{\phi(z)}{(\lambda_k - z)(\lambda_j - z)} dz = \begin{cases} \phi'(\lambda_j) & \text{if } k = j, \\ \frac{\phi(\lambda_k) - \phi(\lambda_j)}{\lambda_k - \lambda_j} & \text{if } k \neq j. \end{cases}$$

Equation (98) can then be written as

$$\phi(\tilde{B}) - \phi(B) = \sum_{j\geq 1} \phi'(\lambda_j) P_j A P_j + \sum_{k\neq j} \frac{\phi(\lambda_k) - \phi(\lambda_j)}{\lambda_k - \lambda_j} P_k A P_j + \mathcal{O}(A^2).$$

Now, since $\phi(\tilde{B}) = \phi(B) + \phi'_B A + \mathcal{O}(A^2)$, the Frechet derivative at B is

$$\phi'_B A = \sum_{j\geq 1} \phi'(\lambda_j) P_j A P_j + \sum_{k\neq j} \frac{\phi(\lambda_k) - \phi(\lambda_j)}{\lambda_k - \lambda_j} P_k A P_j. \tag{99}$$

Equation (99) will be used extensively when we consider the delta method for functions of random operators.

References

[1] T. Anderson, *An Introduction to Multivariate Statistical Analysis: Third Edition*, Wiley, New York, 2003.

[2] N. Aronszajn, *Theory of Reproducing Kernels*, American Mathematical Society Transactions **68** (1950), pp. 337–404.

[3] R. Ash and M. Gardiner, *Topics in Stochastic Processes*, Academic Press, 1975.

[4] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Klewer Academic Publishers, 2004.

[5] P. J. Bickel, E. Levina, *Regularized estimation of large covariance matrices*, The Annals of Statistics **36** (2008) pp. 199227.

[6] J. Cupidon, R. Eubank, D. Gilliam, and F. Ruymgaardt, *Some Properties of Canonical Correlations and Variates in Infinite Dimensions*, Journal of Multivariate Analysis, **99** (2006), pp. 1083–1104.

[7] J. Cupidon, R. Eubank, D. Gilliam, and F. Ruymgaardt, *The Delta-Method for Analytic Functions of Random Operators with Application to Functional Data*, Linear Algebra and its Applications, **13** (2007), pp. 1179–1194.

[8] J. Dauxois, G. M. Nkiet, and Y. Romain, *Canonical Analysis Relative to a Closed Subspace*, Linear Algebra and it's Applications, **388** (2004), pp. 119–145.

[9] J. Dauxois, G. M. Nkiet, and Y. Romain, *Asymptotic Theory for the Principal Component Analysis of a Vector Random Function*, Journal of Multivariate Analysis, **12** (1982), pp. 136–154.

[10] J. Dauxois, A. Pousse, and Y. Romain, *Une extension l'analyse canonique. Quelueqes applications*, Annales de L' I.H.P., Section B, **11** (1975), pp. 355–379.

[11] A.C. Davison and D.V. Hinkley, *Bootstrap Methods and their Application*, Cambridge University Press, New York, 1997.

[12] J. Doob, *Stochastic Processes*, Wiley, New York, 1953.

[13] N. Dunford and J. Schwartz, *Linear Operators: Volumes I, II and III*, Interscience Publishers Inc., New York, 1958.

[14] H. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, 2000.

[15] R. L. Eubank and T. Hsing, *Canonical Correlation for Stochastic Processes*, Stochastic Processes and their Application, **118** (2007), pp. 1634–1661.

[16] E. Hannan, *The General Theory of Canonical Correlations and its Relation to Functional Analysis*, Journal of the Australian Mathematical Society, **2** (1961), pp. 229–242.

[17] G. He, H. Müller, and J. Wang, *Functional Canonical Analysis for Square Integrable Stochastic Processes*, Journal of Multivariate Analysis, **85** (2000), pp. 54–77.

[18] G. He, H. Müller, and J. Wang, *Methods of Canonical Analysis for Functional Data*, Journal of Statistical Planning and Inference, **122** (2002), pp. 141–159.

[19] H. Hotelling, *Relations Between Two Sets of Variates*, Biometrica, **28** (1936), pp. 321–377.

[20] D. King, *A Note on the Cross-Covariance Operator and on Congruence Relations for Hilbert Space Valued Stochastic Processes*, Statistics and Probability Letters, **80** (2010), pp. 361–365.

[20] C. Khatri, *A note on Multiple and Canonical Correlation for a Singular Covariance Matrix*, Psychometrica, **41** (1976), pp. 465–470.

[21] R. Johnson and D. Wichern, *Applied Multivariate Statistical Analysis: Fifth Edition*, Prentise Hall, New Jersey, 2002.

[22] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, New York, 1980.

[23] A. Kupressanin, *Topics in Functional Canonical Correlation and Regression*, Ph.D. Thesis, Dept. Math and Stats., Arizona State University, 2008.

[24] R. G. Laha and V. K. Rohatgi, *Probability Theory*, Wiley, New York, 1979.

[25] S. E. Leurgans., R. A. Moyeed, and B. W. Silverman, *Canonical correlation analysis when the data are curves*, J. Roy. Statist. Soc. B **55** (1993), pp. 725–740.

[26] M. A. Naimark, *Normed Rings*, P. Noordhoff, 1959.

[27] N. H. Nashed and G. Wahba, *Convergence Rates of Approximate Least Squares Solutions to Linear Integral and Operator Equation of the First Kind*, Mathematics of Computation, **28** (1974), pp. 69–80.

[28] E. Parzen, *An Approach to Time Series Analysis*, Ann. Math. Statist., **32** 1961, pp. 951–989.

[29] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer, New York, 2005.

[30] J. O. Ramsay and B. W. Silverman, *Applied Functional Data Analysis*, Springer, New York, 2002.

[31] J. A. Rice and B. W. Silverman, *Estimating the Mean and Covariance Structure Nonparametrically when the Data are Curves*, Statist. Soc. B, **53** (1991), pp. 233-243.

[32] F. Riesz and B. Nagy, *Functional Analysis: 2nd Edition*, Ungar Publishing, 1978.

[33] B. Rynne and M. Youngson, *Linear Functional Analysis*, Springer, London, 2001.

[34] C. R. Vogel, *Computational Method for Inverse Problems*, SIAM, 2002.